1. Elementi di Probabilità e Statistica.

La nascita del calcolo delle probabilità si fa risalire alla seconda metà del '600 e più precisamente al carteggio intervenuto tra Blaise Pascal e Pierre de Fermat a proposito delle questioni poste da Antoine Gombaud Chevalier de Mere.

Il Cavaliere de Mere giocava d'azzardo seguendo la moda dell'epoca e si dice avesse subito gravi perdite scommettendo sull'uscita di almeno una coppia di 6 in 24 lanci di due dadi dopo aver avuto notevoli successi scommettendo sull'uscita di almeno un 6 su 4 lanci di un solo dado. Con le notazioni di oggi posssiamo infatti calcolare che la probabilità di ottenere una coppia di 6 nel lancio di due dadi è $1/36$ e quindi la probabilità di ottenere almeno una coppia di 6 in 24 lanci è

$$1 - \left(\frac{35}{36}\right)^{24} \approx 0.4823$$

essendo $\left(\frac{35}{36}\right)^{24}$ la probabilità di non ottenere una coppia di 6 in nessuno dei 24 lanci

D'altro canto la probabilità di ottenere un 6 nel lancio di un solo dado è $1/6$ e quindi la probabilità di ottenere almeno un 6 in 4 lanci è

$$1 - \left(\frac{5}{6}\right)^{4} \approx 0.5177$$

La questione posta dal cavaliere de Mere era sul tappeto già in precedenza e riguardava la seguente situazione:

Due giocatori A e B scommettono sul successo in almeno 3 *tra* 5 *prove ripetute. Dopo la terza prova A ha ottenuto* 2 *successi e B ne ha ottenuto* 1*. A questo punto si interrompe il gioco. Il problema consiste nel determinare una suddivisione equa della posta*

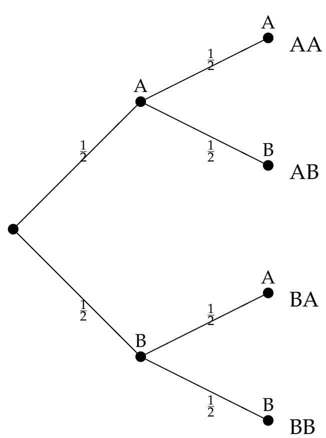
La soluzione di Fermat si basa sul fatto che nel prosieguo del gioco uno solo dei quattro possibili eventi, quello in cui B vinca entrambe le 2 rimanenti partite, è favorevole a B mentre gli altri 3 casi sono tutti favorevoli ad A. Pertanto la posta deve essere divisa nella proporzione di $3/4$ ad A ed $1/4$ a B.

Pascal invece osserva che se il gioco fosse proseguito, poichè nella quarta partita sia A che B hanno eguale possibilità di vittoria, A ha diritto alla metà della posta ed inoltre poichè, se vincesse B, nell'ultima partita le possibilità sarebbero ancora uguali A ha diritto anche alla metà della metà rimanente e quindi in tutto $3/4$ della posta vanno ad A.

Pascal fu anche in grado di ottenere una generalizzazione della sua soluzione estendendo il suo ragionamento per induzione e provando, ad esempio che, nel caso in cui ad A manchino 2 successi e a B ne manchino 3, la posta deve essere divisa in parti proporzionali ai numeri che si ottengono sommando i primi 3 e gli ultimi 2 termini che compaiono nella riga del triangolo aritmetico di Pascal (o Tartaglia) che contiene 5 termini. Il problema era stato affrontato già molte volte nei secoli precedenti ma la soluzione che ora consideriamo corretta fu trovata per la prima volta da Pascal e Fermat e fu formalizzata da Christian Huygens nel suo libro *De ratiociniis in ludo aleae* nel 1657

Nel secolo seguente molti autori pubblicarono libri sull'argomento dando inizio al calcolo delle probabilità. Ricordiamo *Ars conjectandi* di Giacomo Bernoulli del 1713, *Essay d'analyse sur les jeux de hasard* di Pierre Rémond de Montmort pubblicato nel 1708 e nel 1711, *Doctrine of chances* di Abraham De Moivre pubblicato nel 1718, nel 1738 e nel 1756, *Doctrine of annuity and reversions* di Thomas Simpson del 1742, *Annuities on lives* di Abraham De Moivre pubblicato nel 1725, nel 1743, nel 1750 e nel 1752.

Prima ancora possiamo ricordare i contributi precursori della teoria delle probabilità dovuti a Cardano contenuti nel *Liber de ludo aleae*, probabilmente scritti nel 1560 ma, pubblicati postumi dopo l'uscita del lavoro di Huygens.



1.0.1 La divisione della posta

Consideriamo il problema della divisione della posta che abbiamo prima introdotto, quando ai due giocatori, che chiameremo A e B mancano rispettivamente 1 e 2 partite.

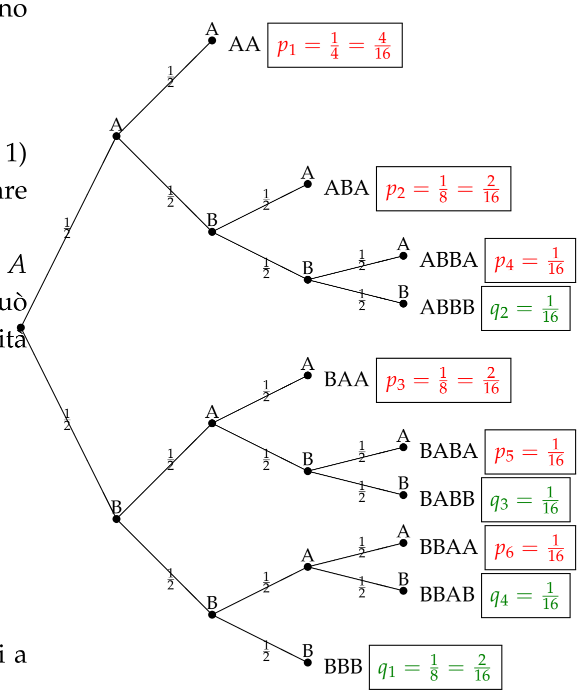
Abbiamo già visto che la posta deve essere divisa, in questo caso, in parti proporzionali a 3 e 1; in altre parole al primo giocatore spetta $\frac{3}{3+1} = \frac{3}{4}$, mentre al secondo spetta $\frac{1}{3+1} = \frac{1}{4}$ della posta. Possiamo ricavare lo stesso risultato usando un semplice grafo ad albero che elenca tutte i possibili esiti di 2 partite, tante quante ne servono per concludere il gioco.

I casi favorevoli ad A sono quindi quelli in cui A compare almeno una volta e sono 3, mentre l'unico caso favorevole a B è quello in cui B compare due volte. È evidente l'analogia con i monomi che compaiono nello sviluppo di

$$(A+B)^2 = A^2 + 2AB + B^2$$

e possiamo anche notare che la somma di tutti i coefficienti $(1+2+1)$ è 4, che la somma dei primi due è 3 mentre il terzo è 1 e congetturare che ci sia una relazione con la suddivisione equa della posta.

Per capire qualcosa in più consideriamo allora il caso in cui ad A manchino 2 partite e a B ne manchino 3. In tal caso la situazione può essere descritta enumerando i casi possibili ed indicando la probabilità di accadimento p_i o q_i di ognuno, come segue.



Pertanto la probabilità di vittoria di A è

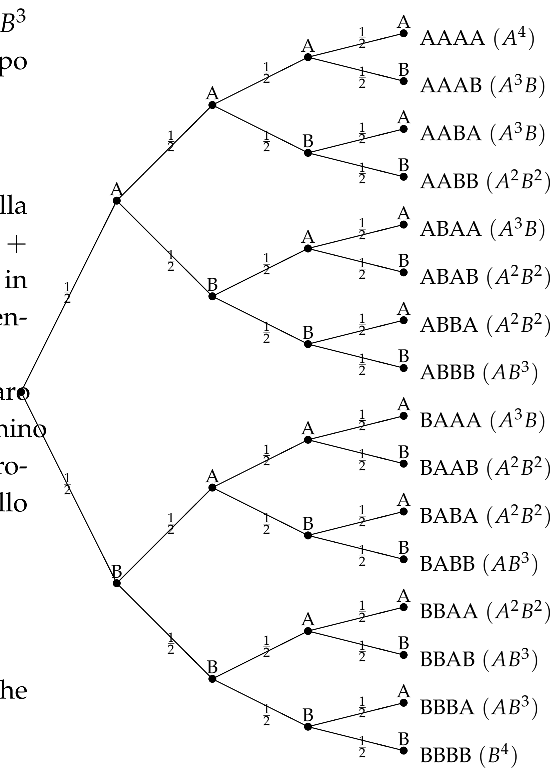
$$p = \sum p_i = \frac{1}{16}(4+2+1+2+1+1) = \frac{11}{16}$$

mentre B vince con probabilità

$$q = \sum q_i = \frac{1}{16}(1+1+2+1) = \frac{5}{16}$$

Ovviamente $p+q=1$ e la posta va divisa in parti proporzionali a 11 e 5

Completiamo il grafo ad albero elencando tutti i casi possibili, ognuno di essi ha uguale probabilità $p_i = \frac{1}{16}$; identifichiamo poi le possibili uscite con monomi in A e B e contiamone il numero.



Osserviamo che A^4 compare 1 volta, A^3B 4 volte, A^2B^2 6 volte, AB^3 4 volte e B^4 1 volta ed è immediato notare l'analogia con lo sviluppo della quarta potenza del binomio

$$(a+b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$$

I casi favorevoli ad A sono quelli in cui a compare almeno alla potenza 2 e possiamo contarli sommando i relativi coefficienti $1+4+6$; in maniera analoga otteniamo che i casi favorevoli a B son quelli in cui b compare almeno alla terza potenza e anche qui possiamo ottenerne il numero sommando i relativi coefficienti $4+1$.

Otteniamo quindi lo stesso risultato già visto ed inoltre risulta chiaro come estendere la regola della suddivisione al caso in cui ad A manchino k vittorie e a B ne manchino h. La suddivisione dovrà essere proporzionale alla somma dei primi k e degli ultimi h coefficienti dello sviluppo di $(a+b)^{k+h-1}$.

1.1 *Qualche richiamo di calcolo combinatorio.*

Per studiare un po' di probabilità discreta è utile conoscere qualche elemento di calcolo combinatorio.

Il calcolo combinatorio si occupa di stabilire il numero delle possibili uscite di semplici esperimenti; si fonda essenzialmente sul principio seguente:

Se un esperimento ha n_1 possibili esiti, un secondo esperimento ha n_2 possibili esiti, un terzo esperimento ha n_3 possibili esiti, allora il numero dei possibili esiti della sequenza dei tre esperimenti è

$$n_1 n_2 n_3$$

Le più comuni conseguenze di questo principio portano a un certo numero di definizioni che descriviamo brevemente.

1.1.1 *Disposizioni di n elementi a k a k.*

Parliamo di **disposizioni** (o anche, se $k = n$, di **permutazioni**) di n elementi a k a k quando consideriamo i gruppi che si ottengono scegliendo k elementi tra gli n dati.

Riteniamo due gruppi distinti se differiscono per un elemento o per l'ordine con cui gli elementi sono scelti.

Indichiamo con

$${}_nD_k$$

il numero delle disposizioni di n elementi a k a k.

Poichè per il primo elemento di ciascun gruppo abbiamo n scelte, per il secondo ne abbiamo $(n-1)$ per il terzo ne abbiamo $(n-2)$ e così via, possiamo calcolare che

$${}_nD_k = n(n-1)(n-2)(n-3)....(n-(k-1)) = \frac{n!}{(n-k)!}$$

Inoltre

Il numero delle disposizioni di n elementi ad n ad n, cioè delle permutazioni, risulta

$$P_n =_n D_n = n!$$

Parliamo di **permutazioni con elementi ripetuti** quando consideriamo le permutazioni di n elementi che si presentano in k sottogruppi di elementi indistinguibili, ciascuno composto da n_1, n_2, n_3,...,n_k elementi con $n_1 + n_2 + n_3 + \cdots + n_k = n$

> Le permutazioni con elementi ripetuti risultano in numero di
> $$\frac{n!}{n_1!n_2!n_3!\cdots n_k!}$$

Nella formula precedente il denominatore è giustificato dal fatto che, per l'$i-$esimo gruppo, ci sono $n_i!$ modi di scegliere in sequenza cli n_i elementi.

Qualora si possa scegliere da n elementi per coprire k posti senza il vincolo di non considerare un elemento giá scelto parliamo di disposizioni con ripetizione.

> Si calcola facilmente che le disposizioni con ripetizione sono in numero di
> $$n^k$$

1.1.2 *Combinazioni di n elementi a k a k.*

Individuare una **combinazione** di n elementi a k a k significa assegnare k elementi ad $n \geq k$ posizioni senza tener conto dell'ordine con cui gli elementi figurano.

> Indichiamo con
> $${}_nC_k$$
> il numero delle combinazioni di n elementi a k a k.

Possiamo visualizzare mediante k segni ✺ gli elementi cui bisogna assegnare una posizione e con n circoletti ○ le posizioni disponibili Ad esempio nella figura sono riportate 22 posizioni e 9 elementi.

○ ⊛ ○ ○ ○ ⊛ ○ ⊛ ⊛ ○ ○ ⊛ ⊛ ○ ○ ⊛ ⊛ ○ ○ ○ ○ ⊛

Naturalmente non è importante in quale delle 9 posizioni occupate da ✺ si colloca il primo elemento, in quale si colloca il secondo, in quale il terzo e così via; quindi ci sono molti modi per disporre sulle 9 posizioni occupate i 9 elementi che indichiamo con

❶ ❷ ❸ ❹ ❺ ❻ ❼ ❽ ❾

La seguente figura descrive una possibile scelta

○ ❶ ○○○ ❷ ○ ❸ ❹ ○○ ❺ ❻ ○○ ❼ ❽ ○○ ○○ ❾

in cui il primo elemento è collocato sulla prima posizione scelta il secondo sulla seconda e così fino al nono.

Altre possibili disposizioni sono
○ ❸ ○ ○ ○ ❷ ○ ❶ ❾ ○ ○ ❺ ❽ ○ ○ ❼ ❻ ○ ○ ○ ○ ❹
oppure
○ ❾ ○ ○ ○ ❷ ○ ❽ ❹ ○ ○ ❺ ❻ ○ ○ ❶ ❸ ○ ○ ○ ○ ❺

Quindi per stimare il numero di possibilità di disporre k segni ✺ in n posizioni ○ occorre

- contare in quanti modi si possono scegliere k posizioni su n disponibili; sia N_k il loro numero.
- contare in quanti modi si possono disporre i k segni ✺ (elementi) sulle k posizioni scelte; sia K_k il loro numero.
- Calcolare $\frac{N_k}{K_k}$ in quanto ogni possibilità di disporre k segni ✺ in n posizioni è ottenuta K_k volte, se non si tiene conto dell'ordinamento.

Avremo, in tutto,

$$N_k = n(n-1)(n-2)(n-3)\cdot\cdot(n-(k-1)) = \frac{n!}{(n-k)!}$$

modi possibili

e

$$K_k = k(k-1)(k-2)(k-3)\cdot\cdot 1 = k!$$

Concludendo, il numero di modi in cui si possono disporre k elementi su n lanci è

$$\frac{n(n-1)(n-2)(n-3)\cdot\cdot(n-(k-1))}{k!} = \frac{n!}{(n-k)!k!} = \binom{n}{k}$$

Poniamo

$$\boxed{{}_nC_k = \binom{n}{k} = \frac{n!}{k!(n-k)!}}$$

Il numero ${}_nC_k$ si chiama coefficiente binomiale.

Ricordiamo che i coefficienti binomiali possono essere ricavati dal triangolo di Tartaglia e che trovano una importante applicazione nella formula del binomio di Newton che illustriamo brevemente di seguito.

Lemma 1.1 *(Triangolo di Tartaglia)*

$$\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}$$

Dimostrazione.

$$\binom{n}{k} + \binom{n}{k-1} = \frac{n!}{k!(n-k)!} + \frac{n!}{(k-1)!(n-(k-1))!} =$$
$$= \frac{n![(n+1-k)+k]}{k!(n+1-k)!} = \frac{(n+1)!}{k!(n+1-k)!} = \binom{n+1}{k}$$

$\square$

Possiamo allora costruire una tabella con le righe indicizzate da n e le colonne indicizzate da k ponendo 1 nei posti corrispondenti a $k = 0$ e $k = n$ e calcoliamo ogni elemento sommando i due elementi della riga precedente, che occupano la stessa colonna e quella immediatamente a sinistra della posizione occupata dall'elemento considerato.

In virtù dell'uguaglianza precedente la tabella contiene nella $k-$esima colonna della $n-$esima riga il coefficiente binomiale $\binom{n}{k}$ e prende il nome di triangolo di Tartaglia, o di Pascal; per il modo semplice e iterativo con cui è costruita, risulta molto comoda per calcolare i coefficienti binomiali.

n \ k	0	1	2	3	4	5	6
1	1	1					
2	1	2	1				
3	1	3	3	1			
4	1	4	6	4	1		
5	1	5	10	10	5	1	
6	1	6	15	20	15	6	1

Table 1.1: Il triangolo di Tartaglia per $n \leq 6$

$$\begin{array}{cccccc}
\binom{1}{0} & \binom{1}{1} & & & & \\
\binom{2}{0} & \binom{2}{1} & \binom{2}{2} & & & \\
\binom{3}{0} & \binom{3}{1} & \binom{3}{2} & \binom{3}{3} & & \\
\dots & \dots & \dots & \dots & \dots & \\
\binom{n}{0} & \dots & \binom{n}{k-1} & \binom{n}{k} & \dots & \binom{n}{n} \\
\dots & \dots & \dots & \binom{n+1}{k} & \dots & \dots \\
\dots & \dots & \dots & \dots & \dots & \dots
\end{array}$$

Valgono inoltre, per i coefficienti binomiali, le seguenti proprietà che risultano molto utili in alcuni calcoli che riguardano le distribuzioni di probabilità discrete.

Lemma 1.2 *Si ha*

$$k\binom{n}{k} = n\binom{n-1}{k-1}$$

Dimostrazione. Basta eseguire il calcolo algebrico. $\square$

Lemma 1.3 *- Identità di Vandermonde - Si ha*

$$\binom{m+n}{k} = \sum_{h=0}^{k}\binom{m}{h}\binom{n}{k-h}$$

Dimostrazione. Osserviamo che

- $\binom{m+n}{k}$ è il numero dei modi con cui si possono scegliere k elementi tra $m+n$
- $\binom{m}{h}$ è il numero dei modi con cui si possono scegliere h elementi tra m
- $\binom{n}{k-h}$ è il numero dei modi con cui si possono scegliere $k-h$ elementi tra n

Poichè si possono scegliere k elementi tra $m+n$ prendendone h tra i primi m e $k-h$ tra gli altri n, possiamo contare in quanti modi questo si può fare semplicemente tenendo conto che, per h fissato ci sono $\binom{m}{h}\binom{n}{k-h}$ possibili scelte.

Sommando su h si trovano tutte e si ottiene la formula. $\square$

Teorema 1.1 *(Binomio di Newton)*

$$(a+b)^n = \sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k$$

Dimostrazione. E' immediato verificare che la formula vale per $n = 1$.

Proviamo ora che, se la formula è valida per n, allora è valida anche per $n+1$. Si ha

$$
\begin{aligned}
&(a+b)^{n+1} = (a+b)^n(a+b) = \\
&= \left(\sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k\right)(a+b) = a\sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k + b\sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k = \\
&= \sum_{k=0}^{n}\binom{n}{k}a^{n+1-k}b^k + \sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^{k+1} = \\
&= a^{n+1} + \sum_{k=1}^{n}\binom{n}{k}a^{n+1-k}b^k + \sum_{k=0}^{n-1}\binom{n}{k}a^{n-k}b^{k+1} + b^{n+1} = \qquad (1.1)
\end{aligned}
$$

$$
\begin{aligned}
&= a^{n+1} + \sum_{k=1}^{n}\binom{n}{k}a^{n+1-k}b^k + \sum_{k=1}^{n}\binom{n}{k-1}a^{n+1-k}b^k + b^{n+1} = \\
&= a^{n+1} + \sum_{k=1}^{n}\left(\binom{n}{k} + \binom{n}{k-1}\right)a^{n+1-k}b^k + b^{n+1} = \\
&= a^{n+1} + \sum_{k=1}^{n}\binom{n+1}{k}a^{n+1-k}b^k + b^{n+1} = \sum_{k=0}^{n+1}\binom{n+1}{k}a^{n+1-k}b^k
\end{aligned}
$$

□

1.1.3 *Campioni ordinati*

Il calcolo combinatorio è utile per stimare il numero di possibili campioni estratti da una popolazione.

Per aiutarci assimiliamo la popolazione ad un un'urna piena di palline e l'estrazione degli elementi del campione all'estrazione delle palline dall'urna.

Possiamo operare un campionamento con ripetizione estraendo una pallina, osservandola e rimettendola nell'urna dopo aver annotato l'informazione relativa.

In tal caso, se operiamo k estrazioni, avremo

$$\overbrace{nnn...n}^{k\ volte} = n^k$$

possibili uscite in quanto per ogni elemento estratto avremo sempre n possibili scelte.

Nel caso in cui si operi invece un campionamento senza ripetizione, estraendo, osservando e non rimettendo la pallina nell'urna, per la prima estrazione avremo n possibilità, per la seconda $n-1$, per la terza $n-3$ e così via.

Pertanto avremo

$${}_nD_k = \frac{n!}{(n-k)!}$$

possibili uscite.

1.2 Spazi di probabilità

1.2.1 Il Lancio di una moneta

Per introdurre i formalismi necessari per parlare di probabilità cominciamo ad illustrare qualche semplice esempio di quello che chiameremo spazio di probabilità discreto Consideriamo il più semplice tra i giochi d'azzardo, e cioè il lancio di una moneta. Possiamo schematizzare il gioco introducendo gli eventi possibili per un singolo lancio, che sono:

- L'uscita di Testa T
- L'uscita di Croce C

È naturale definire la loro probabilità di accadimento ponendo

$$\mathcal{P}(T) = \frac{1}{2} \ , \ \mathcal{P}(C) = \frac{1}{2}$$

e possiamo motivare la nostra scelta con il fatto che le uscite possibili sono due una sola delle quali è considerata per l'evento T o l'evento C

Naturalmente la definizione presuppone che T e C si presentino con ugual frequenza, cioè che la moneta sia non truccata; inoltre va detto che consideriamo una astrazione del gioco in quanto non possiamo escludere che una moneta reale, dopo essere stata lanciata si fermi in una posizione che non corrisponda a nessuna delle due cui attribuiamo il significato di T o C ed inoltre, nella realtà, non è possibile essere certi che la moneta non presenti una faccia più frequentemente di un'altra per causa della sua conformazione.

Accanto agli eventi elementari possiamo introdurre anche l'evento

$$\mathcal{U} = \{\mathbf{T}, \mathbf{C}\}$$

che rappresenta l'evento certo e l'evento $\varnothing$ che assume invece il ruolo di evento impossibile. Chiaramente

$$\mathcal{P}(\mathcal{U}) = \mathcal{P}(T) + \mathcal{P}(C) = 1 \ , \mathcal{P}(\varnothing) = 0$$

Osserviamo che abbiamo quindi definito

- Un insieme $\mathcal{U}$ che contiene tutte le possibili uscite del gioco
- una famiglia di insiemi $\mathcal{F}$ costituita da tutti gli eventi che possiamo considerare
$$\mathcal{F} = \{\mathcal{U} \ , \ T \ , \ C \ , \ \varnothing\}$$
- una funzione $\mathcal{P}$ che associa ad ogni evento un numero positivo con la condizione che $\mathcal{P}(\mathcal{U}) = 1$

1.2.2 *Il lancio di un dado*

Un caso del tutto simile è quello in cui si considera un dado con le facce numerate da 1 a 6; identifichiamo con x_k l'evento *è stato ottenuto il punteggio k* (cioè la faccia superiore del dado mostra k) per $k = 1..6$.

Evidentemente possiamo attribuire ad ogni evento x_k una probabilità tenendo conto che ciascun evento è individuato da uno dei 6 casi possibili. Astraendo anche qui possiamo scrivere che:

Evento	x_1	x_2	x_3	x_4	x_5	x_5
Probabilità	1/6	1/6	1/6	1/6	1/6	1/6

In altre parole

$$P(x_k) = \frac{1}{6} = p_k$$

Chiaramente possiamo anche qui individuare un insieme che contenga tutti gli eventi

$$\mathcal{U} = \{x_1\ ,\ x_2\ .\ x_2\ .\ x_2\ .\ x_2\ .\ x_2\ .\ x_6\ \}$$

per il quale si ha

$$\mathcal{P}(\mathcal{U}) = \sum_{k=1}^{6} \mathcal{P}(x_k) = \sum_{k=1}^{6} \frac{1}{6} = 1$$

e che rappresenta l'evento certo e un insieme $\varnothing$ che rappresenta l'evento impossibile per il quale ovviamente si ha

$$\mathcal{P}(\varnothing) = 0$$

In questo caso possiamo anche considerare molti altri eventi come, ad esempio l'evento E_p che è individuato dall'uscita di un pari:

$$E_p = \{x_2\ ,\ x_4\ ,\ x_6\ \}$$

per cui si ha

$$\mathcal{P}(E_p) = \sum_{k=1}^{3} \mathcal{P}(x_{2k}) = \sum_{k=1}^{3} \frac{1}{6} = \frac{1}{2}$$

oppure

$$E_1 = \{x_2\ ,\ x_5\ \}$$

per cui

$$\mathcal{P}(E_1) = \mathcal{P}(x_2) + \mathcal{P}(x_5) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

In generale possiamo considerare tanti eventi quanti sono i sottoinsiemi che si possono formare utilizzando gli elementi di $\mathcal{U}$.

Nella famiglia $\mathcal{F}$ di tali sottoininsiemi, quella che di solito si chiama famiglia delle parti di $\mathcal{U}$, possiamo definire una funzione $\mathcal{P}$ che assegna ad ogni $E \in \mathcal{F}$ un numero $\mathcal{P}(E)$ che si ottiene semplicemente sommando $\frac{1}{6}$ per tante volte quanti sono gli elementi di E.

Anche in questo caso quindi avremo ottenuto una terna

$$(\mathcal{U}, \mathcal{F}, \mathcal{P})$$

che individua quello che possiamo chiamare uno spazio di probabilità.

1.2.3 Il lancio di due dadi

Consideriamo ora il caso del lancio di due dadi.

Se le facce sono numerate, come al solito, da 1 a 6 possiamo identificare l'esito del lancio con la coppia di numeri (i, j) (punteggio) che si leggono sulla faccia superiore del primo e del secondo dado.

In tal modo possiamo identificare ciascuna delle 36 possibili uscite (eventi) con il punto del piano cartesiano di coordinate (i, j); indicheremo tale evento con il simbolo $A_{i,j}$, (si veda la figura 1.1).

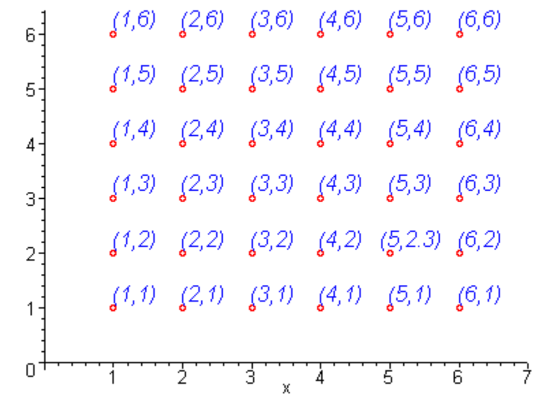


Figure 1.1: Lo spazio $\mathcal{U}$ degli eventi nel caso del lancio di due dadi

Poichè nel caso di dadi non truccati ogni evento è equiprobabile possiamo affermare che la probabilità di $A_{i,j}$ è data da

$$\mathcal{P}(A_{i,j}) = \frac{1}{36}$$

Ovviamente possiamo combinare gli eventi elementari per costruire altri eventi; ad esempio possiamo considerare un nuovo evento

$$B = A_{1,4} \cup A_{5,6} = \{A_{1,4}, A_{5,6}\}$$

e, dal momento che B contiene 2 eventi elementari sui 36 possibili, possiamo ragionevolmente definire

$$\mathcal{P}(B) = \frac{2}{36} = \frac{1}{36} + \frac{1}{36}$$

Anche in questo caso abbiamo quindi costruito una terna

$$(\mathcal{U}, \mathcal{F}, \mathcal{P})$$

che costituisce lo spazio di probabilità che rappresenta il lancio di due dadi.

1.2.4 Lancio di una moneta reiterato fino al successo

Gli esempi finora considerati riguardano casi in cui il numero di eventi possibili è finito. Possiamo anche costruire esempi in cu si considerino una quantità numerabile di eventi possibili.

Supponiamo do lanciare una moneta per cui

$$\mathcal{P}(T) = p \;,\; \mathcal{P}(C) = 1 - p = q$$

e consideriamo gli eventi E_n individuati dalla condizione che "è uscita testa esattamente all' $n-$esimo lancio.

Avremo che

$$\mathcal{P}(E_k) = (1-p)^{k-1}p$$

in quanto è sempre uscita C per $k-1$ lanci ed è uscito T esattamente al k-esimo lancio

Possiamo usare il grafo ad albero in Figura 1.2 per illustrare la situazione.

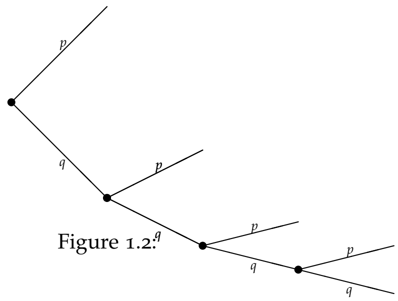


Figure 1.2:

Dal momento che

$$\sum_{1}^{+\infty} p(1-p)^{k-1} = p\sum_{0}^{+\infty}(1-p)^k = p\frac{1}{1-(1-p)} = 1$$

lo spazio costituito da tutti i possibili eventi E_k costituisce uno spazio di probabilità discreto, non finito, numerabile.

1.3 Insiemi e probabilità

Per identificare uno spazio di probabilità discreto possiamo considerare un insieme $\mathcal{U}$, finito o numerabile, che chiameremo spazio dei campioni, i cui elementi $a, b, c \in S$ sono identificabili con gli eventi elementari e i cui sottoinsiemi $A, B, C \subset S$ sono gli eventi.

L'insieme $\mathcal{U}$ sarà quindi identificabile con l'evento certo mentre il vuoto $\varnothing$ sarà l'evento impossibile Possiamo considerare in $\mathcal{U}$ la famiglia $\mathcal{F}$ di tutti i sottoinsiemi di $\mathcal{U}$ ed inoltre gli eventi saranno identificati mediante l'insieme $A \in \mathcal{F}$ sarà identificabile con il fatto che A accade, e il suo complementare A^c con il fatto che A non ha luogo.

$A \cup B$ indicherà che almeno uno tra A e B accade, mentre $A \cap B$ che entrambi A e B accadono.

$A \subset B$ starà ad indicare che se A accade allora necessariamente accade anche B.

$A \cap B = \varnothing$ significherà che A e B non possono accadere simultaneamente e diremo in tal caso che A e B sono mutuamente esclusivi.

Su $\mathcal{F}$ possiamo definire una misura di probabilità semplicemente assegnando una funzione che ad ogni sottoinsieme A di $\mathcal{U}$ assegni un valore $\mathcal{P}(A)$ con le seguenti proprietà:

- Per ogni $A \subset \mathcal{U}$
$$\mathcal{P}(A) \geq 0$$
-
$$\mathcal{P}(\mathcal{U}) = 1$$

- Per ogni famiglia di sottoinsiemi mutuamente esclusivi A_k, $k = 1..n$
$$\mathcal{P}\left(\bigcup_{k=0}^{n} A_k\right) = \sum_{k=0}^{n} \mathcal{P}(A_k)$$

Seguono subito da questi postulati alcuni fatti che possono essere molto utili:

- $\mathcal{P}(\varnothing) = 0$ infatti
$$\mathcal{P}(A) = \mathcal{P}(A \cup \varnothing) = \mathcal{P}(A) + \mathcal{P}(\varnothing)$$
- $\mathcal{P}(A^c) = 1 - \mathcal{P}(A)$ per ogni $A \subset S$
- Se $A \subset B$ allora
$$\mathcal{P}(A) \leq \mathcal{P}(B)$$
infatti
$$\mathcal{P}(B) = \mathcal{P}(A) + \mathcal{P}(B \cap A^c) \geq \mathcal{P}(A)$$
Ne segue che $0 \leq \mathcal{P}(A) \leq \mathcal{P}(S) = 1$ per ogni $A \in \mathcal{F}$
- Se $A \subset B$ allora
$$\mathcal{P}(B \setminus A) = \mathcal{P}(B) - \mathcal{P}(A)$$
infatti
$$\mathcal{P}(B) = \mathcal{P}(A) + \mathcal{P}(B \cap A^c) = \mathcal{P}(B \setminus A)$$
- $\mathcal{P}(A) = \mathcal{P}(A \cap B) + \mathcal{P}(A \cap B^c)$
- $\mathcal{P}\left(A \cup B\right) = \mathcal{P}(A) + \mathcal{P}(B) - \mathcal{P}(A \cap B)$ infatti
$$A \cup B = A \cup (B \setminus (B \cap A))$$
con B e $B \setminus (B \cap A)$ disgiunti, quindi
$$\mathcal{P}\left(A \cup B\right) = \mathcal{P}(A) + \mathcal{P}(B \setminus (B \cap A)) = \mathcal{P}(A) + \mathcal{P}(B) - \mathcal{P}(B \cap A)$$

L'ultima uguaglianza si può generalizzare come

$$\begin{aligned}
&\mathcal{P}(A \cup B \cup C) = \mathcal{P}(A) + \mathcal{P}(B \cup C) - \mathcal{P}(A \cap (B \cup C)) = \\
&= \mathcal{P}(A) + \mathcal{P}(B) + \mathcal{P}(C) - \mathcal{P}(B \cap C) - \mathcal{P}((A \cap B) \cup (A \cap C)) = \\
&= \mathcal{P}(A)+\mathcal{P}(B)+\mathcal{P}(C)-\mathcal{P}(B\cap C)-(\mathcal{P}(A\cap B)+\mathcal{P}(A\cap C)-\mathcal{P}(A\cap B\cap C)) = \\
&= \mathcal{P}(A)+\mathcal{P}(B)+\mathcal{P}(C)-\mathcal{P}(B\cap C)-\mathcal{P}(A\cap B)-\mathcal{P}(A\cap C)+\mathcal{P}(A\cap B\cap C)
\end{aligned}$$

ed anche al caso di più di tre insiemi.

Si ha inoltre che se $\{B_i, i = 1..n\}$ con $B_i \cap B_j = \emptyset$, e $A \subset \bigcup B_i$, allora

$$\mathcal{P}(A) = \mathcal{P}(A \cap B_1) + \mathcal{P}(A \cap B_2) + ... + \mathcal{P}(A \cap B_n) \tag{1.2}$$

In particolare possiamo affermare che

$$A = (A \cap B) \cup (A \cap B^c)$$

e quindi

$$\mathcal{P}(A) = \mathcal{P}(A \cap B) + \mathcal{P}(A \cap B^c)$$

1.3.1 Probabilità condizionata

Definizione 1.1 *Se $A, B \in \mathcal{F}$ definiamo probabilità di A condizionata a B e la denotiamo con $\mathcal{P}(A|B)$ il valore*

$$\mathcal{P}(A|B) = \frac{\mathcal{P}(A \cap B)}{\mathcal{P}(B)}$$

$\mathcal{P}(A|B)$ è la probabilità che A accada nel caso in cui sia accaduto B

Naturalmente si ha

$$\mathcal{P}(A \cap B) = \mathcal{P}(A|B)\mathcal{P}(B)$$

Nel caso in cui

$$\mathcal{P}(A|B) = P(A)$$

diciamo che A e B sono eventi indipendenti, (la probabilità di accadimento di A non è cambiata dal fatto che B è accaduto).

In tal caso si ha

$$\mathcal{P}(A \cap B) = \mathcal{P}(A)\mathcal{P}(B)$$

Vale il seguente

Teorema 1.2 *Sia B_i per $i = 1..n$ una famiglia di eventi in S tali che $B_i \cap B_j = \emptyset$, e $S = \bigcup B_i$, cioè supponiamo che gli insiemi B_i siano mutuamente esclusivi ed esaustivi, allora*

$$\mathcal{P}(A) = \mathcal{P}(A|B_1)\mathcal{P}(B_1) + \mathcal{P}(A|B_2)\mathcal{P}(B_2)) + ...\mathcal{P}(A|B_n)\mathcal{P}(B_n) \tag{1.3}$$

La verifica del teorema segue immediatamente dalla definizione di probabilità condizionata e dalla 1.2.

Si può dimostrare che

Teorema 1.3 *Se $A_i \cap A_j = \varnothing$, ed $A \subset \cup_i A_i$, allora*

$$\mathcal{P}(A) = \mathcal{P}(A_1)\mathcal{P}(A|A_1) + \mathcal{P}(A_2)\mathcal{P}(A|A_2) + ...\mathcal{P}(A_N)\mathcal{P}(A|A_N)$$

infatti

$$\begin{aligned}\mathcal{P}(A) &= \mathcal{P}(A \cap A_1) + \mathcal{P}(A \cap A_2) + ... + \mathcal{P}(A \cap A_N) = \\ &= \mathcal{P}(A_1)\mathcal{P}(A|A_1) + \mathcal{P}(A_2)\mathcal{P}(A|A_2) + ... + \mathcal{P}(A_N)\mathcal{P}(A|A_N)\end{aligned}$$

Da questa semplice considerazione segue facilmente il teorema di Bayes

Teorema 1.4 *- di Bayes - Se $A_1, A_2, ..., A_N \in \mathcal{F}$ sono eventi tali che $A_i \cap A_j = \varnothing$ e $A \subset \bigcup_{i=1}^N A_i$ allora*

$$\mathcal{P}(A_k|B) = \frac{\mathcal{P}(A_k)\mathcal{P}(A|A_k)}{\sum_{i=1}^N \mathcal{P}(A_i)\mathcal{P}(A|A_i)}$$

La dimostrazione del teorema è molto semplice ed è sufficiente vederla nel caso in cui $N = 2$ per comprenderne il meccanismo.

Consideriamo

- $A_1 \cup A_2 \supset A$
- $A_1 \cap A_2 = \varnothing$

si ha

$$\mathcal{P}(A_{1,2}|A) = \frac{\mathcal{P}(A_{1,2} \cap A)}{\mathcal{P}(A)} \qquad \text{e} \qquad \mathcal{P}(A|A_{1,2}) = \frac{\mathcal{P}(A \cap A_{1,2})}{\mathcal{P}(A_{1,2})}$$

ed inoltre

$$\mathcal{P}(A) = \mathcal{P}(A \cap (A_1 \cup A_2)) = \mathcal{P}(A \cap A_1) + \mathcal{P}(A \cap A_2)$$

da cui

$$\mathcal{P}(A_1|A)\mathcal{P}(A) = \mathcal{P}(A \cap A_1) = \mathcal{P}(A|A_1)\mathcal{P}(A_1)$$

Ne segue che

$$\mathcal{P}(A_{1,2}|A) = \frac{\mathcal{P}(A|A_{1,2})}{\mathcal{P}(A)} = \frac{\mathcal{P}(A|A_{1,2})\mathcal{P}(A_{1,2})}{\mathcal{P}(A|A_1)\mathcal{P}(A_1) + \mathcal{P}(A|A_2)\mathcal{P}(A_2)}$$

1.3.2 *Ancora sul lancio di due dadi*

Giocando a dadi è d'uso sommare i punti usciti sull'uno e sull'altro dei due dadi; se chiamiamo ξ il punteggio così ottenuto possiamo allora scrivere che

$$\xi(D_{i,j}) = i + j$$

In questo modo definiamo una funzione ξ su ogni insieme elementare e quindi risulta che

$$\xi : \mathcal{U} \to \mathbb{R}$$

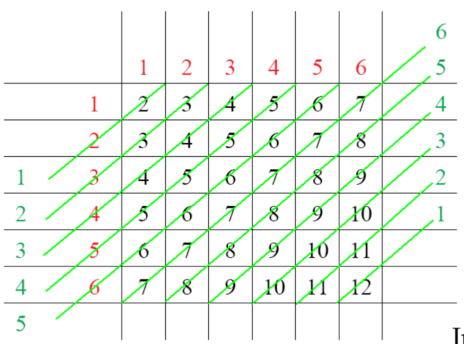
poichè l'uscita di $D_{i,j}$ è casuale, anche $\xi(D_{i,j})$ lo sarà. Se teniamo conto che il punteggio di 3 si può ottenere soltanto in corrispondenza di uno dei due eventi

$$D_{1,2} \text{ e } D_{2,1}$$

scopriamo che la probabilità che ξ assuma il valore 3 è $2/36$ in quanto il valore 3 compare esattamente 2 volte sui 36 casi possibili, in altre parole

$$\mathcal{P}(\xi = 3) = \frac{2}{36}$$

È facile immaginare come calcolare la probabilità che ξ assuma uno dei valori (interi da 2 a 12).



...a figura seguente sono riportate le possibili uscite del lancio di ...di, cioè è rappresentato lo spazio $\mathcal{U}$; Le linee diagonali aiutano a ... quante volte compare ognuno dei valori assunti dalla variabile ... $= i + j$ ed è immediato costruire una tabella in cui siano riassunti ... che ξ può assumere ed il numero di volte, cioè la frequenza, ... compaiono.

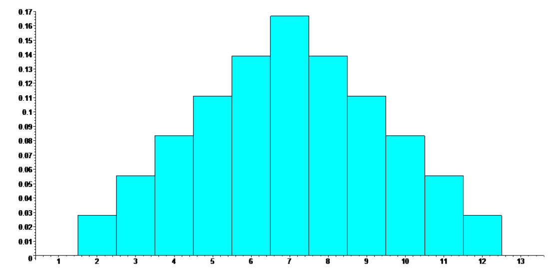
$i+j$	2	3	4	5	6	7	8	9	10	11	12
freq.	1	2	3	4	5	6	5	4	3	2	1

In base alla tabella è immediato ottenere, tenendo anche conto che ognuna delle coppie (i, j) è equiprobabile, che si ha

$$\mathcal{P}(\xi = 2) = \mathcal{P}(\xi = 12) = \frac{1}{36}$$
$$\mathcal{P}(\xi = 3) = \mathcal{P}(\xi = 11) = \frac{2}{36}$$
$$\mathcal{P}(\xi = 4) = \mathcal{P}(\xi = 10) = \frac{3}{36}$$
$$\mathcal{P}(\xi = 5) = \mathcal{P}(\xi = 9) = \frac{4}{36}$$
$$\mathcal{P}(\xi = 6) = \mathcal{P}(\xi = 8) = \frac{5}{36}$$
$$\mathcal{P}(\xi = 7) = \frac{6}{36}$$

e ovviamente
$$\sum_{k=2}^{12} \mathcal{P}(\xi = k) = 1.$$

I risultati possono essere riportati su un istogramma, su un grafico cioè in cui in corrispondenza di ciascun intero k tra 2 e 12 è riportato un rettangolo la cui base $[k - 0.5, k + 0.5]$ ha lunghezza 1 e la cui altezza è pari a $\mathcal{P}(\xi = k)$.



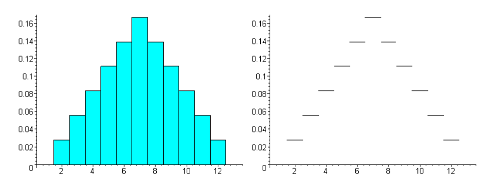
Questo accorgimento consente di valutare la probabilità che $\xi = k$ semplicemente considerando l'area del rettangolo corrispondente; la somma delle aree di tutti i rettangoli sarà ovviamente 1.

Aiutandoci con l'istogramma possiamo facilmente calcolare ad esempio che

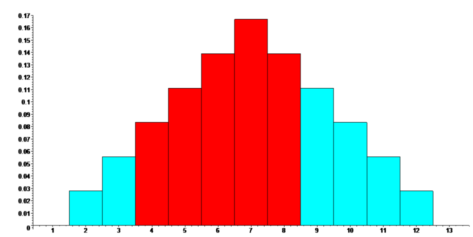
$$P(\xi = 7) = \frac{6}{36} = \frac{1}{6}$$

$$P(\xi \neq 7) = 1 - P(\xi = 7) = 1 - \frac{1}{6} = \frac{5}{6}$$

$$P(4 \leq \xi \leq 8) = P(\xi = 4) + P(\xi = 5) + P(\xi = 6) + P(\xi = 7) + P(\xi = 8) = \frac{3+4+5+6+5}{36} = \frac{23}{36}$$

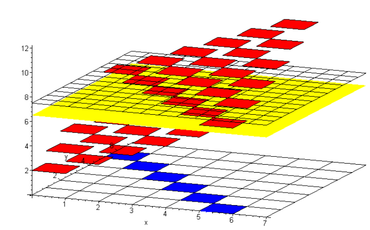


Abbiamo così introdotto un esempio di variabile aleatoria cioè di funzione definita su $\mathcal{U}$ e usando l'istogramma che abbiamo costruito possiamo anche definire il concetto di funzione densità di probabilità (la indicheremo PDF) di una variabile aleatoria. Sarà infatti sufficiente considerare la funzione costante a tratti definita uguale a $\mathcal{P}(\xi = k)$ su $[k - 0.5, k + 0.5]$.



Per capire meglio come si ottiene la funzione densità di probabilità di ξ consideriamo il suo grafico che rappresentiamo assumendo $\xi(i, j) = i + j$ costante sul quadrato $[i - 0.5, i + 0.5] \times [j - 0.5, j + 0.5]$

Ad esempio si vede che $\mathcal{P}(\xi = 7)$ è la probabilità calcolata in $\mathcal{U}$ della controimmagine di 7 secondo ξ cioè $(\xi^{-1}(7)$ o, dal momento che ξ assume valori discreti $(\xi^{-1}((6.5, 7.5))$.



1.3.3 *Spazi di probabilità e variabili aleatorie discrete*

Per assegnare uno spazio di probabilità discreto, finito o numerabile, basta quindi assegnare una famiglia di eventi elementari distinti e disgiunti
$$\mathcal{A} = \{A_i \; : \; i \in \mathcal{I}\}$$

dove $\mathcal{I}$ è un insieme finito o numerabile di indici ed una funzione

$$\mathcal{P} : \mathcal{F} \to \mathbb{R}$$

dove $\mathcal{F}$ è la collezione di tutti i sottoinsiemi (le parti) di $\mathcal{A}$, che associa ad ogni $A \in \mathcal{F}$ un valore reale $\mathcal{P}(A)$, che chiamiamo probabilità che l'evento accada, soddisfacente le seguenti proprietà:

- $\mathcal{P}(A) \geq 0$
- se $\mathcal{U} = \bigcup_{i \in \mathcal{I}} A_i$, si ha $\mathcal{P}(\mathcal{U}) = 1$

  Se $A \in \mathcal{F}$ allora
$$A = \bigcup_{i \in \mathcal{J}} A_i$$
e
$$\mathcal{P}(A) = \sum_{i \in \mathcal{J}} \mathcal{P}(A_i)$$
la somma essendo finita o numerabile.

Poniamo, per semplicità
$$\mathcal{P}(A_i) = p_i$$
ed osserviamo che è sufficiente assegnare $\mathcal{P}(A_i)$ per definire $\mathcal{P}$ su $\mathcal{F}$.

Ci riferiremo quindi allo spazio di probabilità discreto costituito da
$$(\mathcal{U}, \mathcal{F}, \mathcal{P})$$
dove $\mathcal{U}$ è un insieme discreto $\mathcal{F}$ è la famiglia delle parti di $\mathcal{U}$ e $\mathcal{P}$ è una misura di probabilità su $\mathcal{F}$.

Diciamo che è assegnata una variabile aleatoria ξ su $\mathcal{U}$, se è data una funzione
$$\xi : \mathcal{U} \to \mathbb{R}$$
ed indichiamo per brevità
$$\xi_i = \xi(A_i)$$

Definizione 1.2 *Definiamo*

- *la* ***media*** μ *di* ξ *come*
$$\mu = E(\xi) = \sum_i \xi_i p_i$$
- *la* ***varianza*** σ^2 *di* ξ *come*
$$\sigma^2 = \mathrm{Var}(\xi) = E((\xi - \mu)^2) = \sum_i (\xi_i - \mu)^2 p_i$$
- *lo* ***scarto quadratico medio*** *o deviazione standard di* ξ
$$\sigma = \sqrt{\sigma^2}$$
- *il* ***momento*** k***-esimo*** μ_k *di* ξ *come*
$$\mu_k = \sum_i (\xi_i - \mu)^k p_i$$

- *il **momento** k**-esimo rispetto all'origine** μ'_k di ξ come*

$$\mu'_k = \sum_i \xi_i^k p_i$$

- *la **funzione di distribuzione** φ di ξ come*

$$\varphi(\xi_i) = p_i$$

  per cui la funzione di distribuzione cumulativa Φ è definita da:

$$\Phi(x) = \mathcal{P}(\xi \leq x) = \sum_{\xi_i \leq x} \mathcal{P}(\xi(A_i)) = \sum_{\xi_i \leq x} p_i = \sum_i \varphi(\xi_i)$$

Si può dimostrare che, se ξ e η sono variabili aleatorie discrete su uno spazio di probabilità $(\mathcal{U}, \mathcal{F}, \mathcal{P})$ e se $\alpha, \beta \in \mathbb{R}$, allora

- $E(\alpha\xi + \beta\eta) = \alpha E(\xi) + \beta E(\eta)$
- se ξ e η sono variabili aleatorie indipendenti

$$E(\xi\eta) = E(\xi)E(\eta)$$

- $\mathrm{Var}(\alpha\xi) = \alpha^2 \,\mathrm{Var}(\xi)$
- se ξ e η sono variabili aleatorie indipendenti

$$\mathrm{Var}(\xi \pm \eta) = \mathrm{Var}(\xi) + \mathrm{Var}(\eta)$$

e che la varianza si può calcolare come

$$\begin{aligned}\sigma^2 = \mathrm{Var}(\xi) = E((\xi - \mu)^2) = E(\xi^2 - 2\mu\xi + \mu^2) = \\ = E(\xi^2) - 2\mu E(\xi) + \mu^2 = E(\xi^2) - 2\mu^2 + \mu^2 = \\ = E(\xi^2) - \mu^2 = E(\xi^2) - (E(\xi))^2\end{aligned}$$

Inoltre se ξ è una variabile aleatoria discreta la cui densità di probabilità è φ, e se $f : \mathbb{R} \to \mathbb{R}$ è una funzione tale che, posto

$$\eta_i = f(\xi_i)$$

si definisca una corrispondenza biunivoca; η risulta essere una variabile aleatoria la cui densità di probabilità è

$$\psi(k) = \mathcal{P}(\eta = k) = \mathcal{P}(f(\xi) = k) = \mathcal{P}(\xi = f^{-1}(k)) = \varphi(f^{-1}(k))$$

Quindi

$$E(f(\xi)) = E(\eta) = \sum_k k\psi(k) = \sum_i f(\xi_i)\varphi(f^{-1}(f(\xi_i)))) = \sum_i f(\xi_i)\varphi(\xi_i)$$

Definiamo **funzione generatrice dei momenti di ξ** la

$$M_\xi(t) = E(e^{t\xi}) = \sum_i e^{\xi_i t}\varphi(\xi_i)$$

Si può verificare che la funzione M_ξ è sviluppabile in serie di McLaurin ed il suo sviluppo è dato da

$$M_\xi(t) = \sum_{k=0}^{+\infty} \mu'_k \frac{t^k}{k!}$$

per cui

$$\mu'_k = \frac{d^k}{dt^k} M_\xi(t)$$

1.4 Variabili aleatorie continue

Talvolta non è possibile considerare uno spazio di probabilità discreto, finito o numerabile.

Ciò accade, ad esempio, quando si considera il problema di scegliere un numero a caso compreso tra 0 ed 1.

Infatti la probabilità di estrarre, ad esempio, il valore 0.3 non si può calcolare considerando il rapporto tra casi favorevoli, uno solo, e casi possibili, infiniti non numerabili.

Anche la definizione di media e varianza presentano qualche problema in quanto occorre definire come si intende procedere per calcolare la somma di un numero infinito, non numerabile, di addendi.

Per chiarire la questione possiamo osservare che, se è difficile definire la probabilità che la variabile aleatoria ξ il cui valore è il numero scelto a caso in $[0, 1]$ assuma il valore x, è invece naturale definire la probabilità che $\xi \in [x, x + h]$.

In tal caso infatti possiamo identificare i casi favorevoli con un segmento di lunghezza h e la totalità dei casi con l'intero intervallo $[0, 1]$ che risulta ovviamente di lunghezza 1.

Pertanto

$$\mathcal{P}(x \leq \xi \leq x + h) = \frac{h}{1}$$

Ricordando il significato di somma dell'integrale, possiamo definire la funzione distribuzione di probabilità della variabile aleatoria ξ come la funzione continua φ tale che

$$\mathcal{P}(x \leq \xi \leq x + h) = h = \int_\xi^{\xi+h} \varphi(t)dt$$

per ogni $x \in [0, 1]$ e per ogni h abbastanza piccolo.

Ne deduciamo che

$$\frac{1}{h}\int_x^{x+h} \varphi(t)dt = 1$$

e, passando al limite per $h \to 0$, poichè abbiamo supposto φ continua,

$$\varphi(x) = 1$$

Da quanto abbiamo detto appare ragionevole che, nel caso di una variabile aleatoria continua ξ, non è significativo definire

$$\mathcal{P}(\xi = x)$$

mentre è naturale definire

$$\mathcal{P}(x_0 \leq \xi \leq x_1) = \int_{x_0}^{x_1} \varphi(t)dt$$

dove φ è la funzione di distribuzione di probabilità di ξ.

Pertanto supporremo nota una variabile aleatoria continua ξ se è nota la sua funzione di distribuzione di probabilità φ.

Una funzione $\varphi : \mathbb{R} \to \mathbb{R}$, continua, è la funzione di distribuzione di probabilità di una variabile aleatoria se

- $$\varphi(t) \geq 0 \qquad \text{per ogni} t \in \mathbb{R}$$
- $$\int_{-\infty}^{+\infty} \varphi(t)dt = 1$$

In tal caso si ha

$$\mathcal{P}(\xi \leq x) = \int_{-\infty}^{x} \varphi(t)dt \qquad , \qquad \mathcal{P}(x_0 \leq \xi \leq x_1) = \int_{x_0}^{x_1} \varphi(t)dt$$

La funzione

$$F(x) = \mathcal{P}(\xi \leq x) = \int_{-\infty}^{x} \varphi(t)dt$$

si chiama distribuzione cumulativa di probabilità della variabile aleatoria ξ di densità di probabilità φ.

Osserviamo che ad ogni variabile aleatoria discreta (finita) si può associare una variabile aleatoria continua la cui densità è una funzione costante a tratti, nulla al di fuori di un insieme limitato nel caso in cui la variabile sia discreta e finita.

Come nel caso delle variabili aleatorie discrete possiamo porre la seguente

Definizione 1.3 *Se ξ è una variabile aleatoria continua che ha densità di probabilità φ,*

- *la **media** μ di ξ è definita da*

$$\mu = E(\xi) = \int_{-\infty}^{+\infty} x\varphi(x)dx$$

- *la* ***varianza*** *σ^2 di ξ è definita da*

$$\sigma^2 = \mathrm{Var}(\xi) = E((\xi - \mu)^2) = \int_{-\infty}^{+\infty} (x-\mu)^2 \varphi(x) dx$$

- *lo* ***scarto quadratico medio*** *di ξ è definito da*

$$\sigma = \sqrt{\sigma^2}$$

- *la* ***moda*** *M di ξ è definita da*

$$M = \sup_{x \in \mathbb{R}} \varphi(x)$$

- *la* ***mediana*** *m di ξ è definita da*

$$\mathcal{P}(\xi \le m) = \int_{-\infty}^{m} \varphi(x) dx = \int_{m}^{+\infty} \varphi(x) dx = \mathcal{P}(\xi \ge m)$$

- *il* ***momento di ordine*** *k μ_k di ξ è definito da*

$$\mu_k = E((\xi - \mu)^k) = \int_{-\infty}^{+\infty} (x-\mu)^k \varphi(x) dx$$

- *il* ***momento di ordine*** *k****, rispetto all'origine*** *μ'_k di ξ è definito da*

$$\mu_k = E(\xi^k) = \int_{-\infty}^{+\infty} x^k \varphi(x) dx$$

Se ξ è una variabile aleatoria continua e se $f : \mathbb{R} \to \mathbb{R}$ è una funzione derivabile ed invertibile possiamo considerare la variabile aleatoria $f(\xi)$ e possiamo calcolare che

$$\mathcal{P}(x_0 \le f(\xi) \le x_1) = \mathcal{P}(f^{-1}(x_0) \le \xi \le (f^{-1}(x_1)) = \\ \int_{f^{-1}(x_0)}^{f^{-1}(x_1)} \varphi(t) dt = \int_{x_0}^{x_1} \frac{\varphi(f^{-1}(s))}{f'(f^{-1}(s))} ds$$

per cui la sua funzione distribuzione di probabilità risulta definita da

$$\psi(t) = \frac{\varphi(f^{-1}(s))}{f'(f^{-1}(s))}$$

In tal modo si ha

$$E(f(\xi)) = \int_{-\infty}^{+\infty} s \frac{\varphi(f^{-1}(s))}{f'(f^{-1}(s))} ds = \\ = \int_{-\infty}^{+\infty} f(t) \frac{\varphi(t)}{f'(t)} f'(t) dt = \int_{-\infty}^{+\infty} f(t) \varphi(t) dt = \mu$$

ed inoltre

$$\sigma^2(f(\xi)) = \int_{-\infty}^{+\infty} (s-\mu)^2 \frac{\varphi(f^{-1}(s))}{f'(f^{-1}(s))} ds = \\ = \int_{-\infty}^{+\infty} (f(t)-\mu)^2 \frac{\varphi(t)}{f'(t)} f'(t) dt = \int_{-\infty}^{+\infty} (f(t)-\mu)^2 \varphi(t) dt$$

Definizione 1.4 *Se ξ è una variabile aleatoria continua la cui densità di probabilità è φ, definiamo* ***funzione generatrice dei momenti di*** ξ *la*

$$M_\xi(t) = E(e^{t\xi}) = \int_{-\infty}^{+\infty} e^{tx}\varphi(x)dx$$

Possiamo anche in questo caso provare che

- $E(\alpha\xi + \beta\eta) = \alpha E(\xi) + \beta E(\eta)$
- se ξ e η sono variabili aleatorie indipendenti
$$E(\xi\eta) = E(\xi)E(\eta)$$
- $\mathrm{Var}(\alpha\xi) = \alpha^2 \,\mathrm{Var}(\xi)$
- se ξ e η sono variabili aleatorie indipendenti
$$\mathrm{Var}(\xi \pm \eta) = \mathrm{Var}(\xi) + \mathrm{Var}(\eta)$$

Ed è utile ricordare ancora che

$$\begin{aligned}\sigma^2 = E((\xi-\mu)^2) = E(\xi^2 - 2\mu\xi + \mu^2) = \\ = E(\xi^2) - 2\mu E(\xi) + \mu^2 = \\ = E(\xi^2) - 2\mu^2 + \mu^2 = E(\xi^2) - \mu^2 = E(\xi^2) - (E(\xi))^2\end{aligned}$$

Si ha inoltre:

$$(t-\mu)^k = \sum_{i=0}^{k} \binom{k}{i} t^i \mu^{k-i}$$

moltiplicando per $\varphi(t)$ ed integrando,otteniamo

$$\mu_k = \sum_{i=0}^{k} \binom{k}{i} {\mu'}_i \mu^{k-i} \qquad (1.4)$$

e se ne ricava che per trovare i momenti rispetto al valor medio μ_k è sufficiente conoscere i momenti rispetto all'origine μ'_k.

Casi particolari della 1.4 sono

$$\mu_2 = \sigma_2 = \mu'_2 - \mu^2$$
$$\mu_3 = \mu'_3 - 3\mu'_2\mu + 2\mu^3$$

(ricordiamo che $\mu_0 = \mu'_0 = 1$ e $\mu_1 = \mu'_1 = \mu$).

La funzione generatrice dei momenti si rivela molto comoda per il calcolo dei momenti di una variabile aleatoria.

Infatti si può verificare che M_ξ è sviluppabile in serie di McLaurin ed il suo sviluppo è dato da

$$M_\xi(t) = \int_{-\infty}^{+\infty} \sum_{i=0}^{+\infty} \frac{(st)^k}{k!} \varphi(s) ds = \sum_{i=0}^{+\infty} \int_{-\infty}^{+\infty} s^k \varphi(s) ds \frac{t^k}{k!}$$

per cui

$$M_\xi(t) = \sum_{i=0}^{+\infty} \mu'_k \frac{t^k}{k!}$$

e quindi

$$\mu'_k = \frac{d^k}{dt^k} M_\xi(t)$$

1.5 La disuguaglianza di Tchebichev e la legge dei grandi numeri

In questa sezione ci occupiamo di due risultati fondamentali: la disuguaglianza di Tchebichev e la legge dei grandi numeri, cominciando a parlare della prima.

Sia ξ una variabile aleatoria con media μ e varianza σ^2, allora si avrà che

$$\begin{aligned}
\sigma^2 = \int_{-\infty}^{+\infty} (t-\mu)^2 \varphi(t) dt = \\
= \int_{\{t \,:\, |t-\mu| \geq \varepsilon\}} (t-\mu)^2 \varphi(t) dt + \int_{\{t \,:\, |t-\mu| < \varepsilon\}} (t-\mu)^2 \varphi(t) dt \geq \\
\geq \int_{\{t \,:\, |t-\mu| \geq \varepsilon\}} (t-\mu)^2 \varphi(t) dt \geq \int_{\{t \,:\, |t-\mu| \geq \varepsilon\}} \varepsilon^2 \varphi(t) dt = \\
= \varepsilon^2 \mathcal{P}(|\xi - \mu| \geq \varepsilon)
\end{aligned}$$

se ne ricava pertanto che

$$\mathcal{P}(|\xi - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \tag{1.5}$$

La 1.5 è nota come **disuguaglianza di Tchebichev** e ne possiamo trarre una interessante conseguenza: per $\varepsilon = k\sigma$ otteniamo che

$$\mathcal{P}(|\xi - \mu| \geq k\sigma) \leq \frac{1}{k^2} \tag{1.6}$$

Pertanto

$$\mathcal{P}(|\xi - \mu| < k\sigma) = 1 - \mathcal{P}(|\xi - \mu| \geq k\sigma) \geq 1 - \frac{1}{k^2} \tag{1.7}$$

Se ora consideriamo la seguente tabella

k	1	2	3	4	5	6
$1-\frac{1}{k^2}$	0	.75	.88	.93	.95	.97

Table 1.2: Valori approssimati di $1-\frac{1}{k^2}$

Si vede pertanto che se ξ è una variabile aleatoria di media μ e di varianza σ^2, allora la probabilità che il valore assunto da ξ sia vicino alla media μ per meno di 2 volte la varianza è del 75% e sale all'88% se ci accontentiamo di un errore inferiore a 3 volte la varianza.

Va osservato che, nonostante fornisca risultati soddisfacenti, la disuguaglianza di Tchebichev non è molto precisa.

1.5.1 *La legge dei grandi numeri*

Una delle conseguenze della disuguaglianza di Tchebichev prende il nome di "Legge dei Grandi Numeri" e si ricava come segue.

Se $\xi_1, \xi_2, \ldots, \xi_n$ sono variabili aleatorie tutte con media μ e varianza σ^2, la variabile aleatoria

$$S_n = \frac{\xi_1 + \xi_2 + \cdots + \xi_n}{n}$$

ha media

$$E(S_n) = \frac{1}{n}(E(\xi_1) + E(\xi_2) + \cdots + E(\xi_n)) = \mu$$

e varianza

$$\mathrm{Var}(S_n) = \frac{1}{n^2}(\mathrm{Var}(\xi_1) + \mathrm{Var}(\xi_2) + \cdots + \mathrm{Var}(\xi_n)) = \frac{\sigma^2}{n}$$

inoltre vale il seguente teorema

Teorema 1.5 *Siano $\xi_1, \xi_2, \ldots, \xi_n$ variabili aleatorie tutte con media μ e varianza σ^2, e consideriamo la variabile aleatoria*

$$S_n = \frac{\xi_1 + \xi_2 + \cdots + \xi_n}{n}$$

Allora

$$\mathcal{P}(|S_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \to 0 \tag{1.8}$$

per $n \to +\infty$

La 1.9 è nota con il nome di "**Legge Debole dei Grandi Numeri**" ed esprime un concetto in base al quale la media di n uscite di una variabile aleatoria differisce dalla media della variabile aleatoria per una quantità infinitesima con n.

Va sottolineato che la legge dei grandi numeri fornisce informazioni di carattere qualitativo e quindi non può essere usata per stime di tipo quantitativo.

È possibile anche dimostrare, ma la dimostrazione è più complessa, la "**Legge Forte dei Grandi Numeri**" che asserisce che

Teorema 1.6 *Siano $\xi_1, \xi_2, \ldots, \xi_n$ variabili aleatorie tutte con media μ e varianza σ^2, e consideriamo la variabile aleatoria*

$$S_n = \frac{\xi_1 + \xi_2 + \cdots + \xi_n}{n}$$

Allora

$$\mathcal{P}(\lim_n S_n = \mu) = 1 \tag{1.9}$$

In entrambi i casi il concetto espresso è che la media di S_n converge alla media μ la differenza risiede nel modo in cui tale convergenza avviene e nelle proprietà che tale convergenza consente di trasferire sul limite. Piú precisamente la legge debole dei grandi numeri afferma che la successione di variabili aleatorie S_n converge a μ in probabilitá, mentre la legge forte dei grandi numeri garantisce che S_n converge a μ quasi certamente.

Usando la terminologia derivante dalla teoria della misura, cui la teoria della probabilità astratta sostanzialmente si sovrappone, la formulazione debole parla di convergenza in misura, mentre la formulazione forte parla di convergenza puntuale quasi ovunque.

È noto che, essendo lo spazio di probabilità di misura finita (uguale ad 1), una successione quasi ovunque convergente è anche convergente in misura, e che da una successione convergente in misura si può estrarre una sottosuccessione quasi ovunque convergente.

1.6 *Somma di variabili aleatorie.*

Consideriamo due variabili aleatorie discrete indipendenti ξ, η aventi PDF rispettivamente f e g, e la variabile aleatoria ζ che restituisce la somma delle due

$$\zeta = \xi + \eta$$

Possiamo trovare la densità di probabilità della variabile ζ osservando che

$$\begin{aligned}
&h(\gamma) = \mathcal{P}(\zeta = \gamma) = \\
&= \sum_\alpha \mathcal{P}(\xi + \eta = \gamma | \xi = \alpha)\mathcal{P}(\xi = \alpha) = \sum_\alpha \mathcal{P}(\eta = \gamma - \alpha)\mathcal{P}(\xi = \alpha) = \\
&\qquad\qquad\qquad\qquad \sum_\alpha f(\gamma - \alpha) g(\alpha)
\end{aligned}$$

Nel caso in cui ξ e η siano variabili aleatorie continue indipendenti

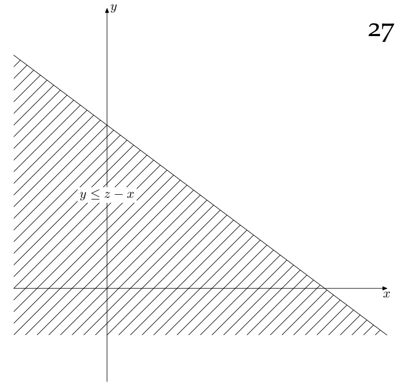


Figure 1.3:

avremo, come si vede dalla figura 1.3,

$$\mathcal{P}(\zeta \leq z) = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z-x} g(y)dy \right) f(x)dx =$$
$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{z} g(s-x)f(x)dsdx = \int_{-\infty}^{z} \left(\int_{-\infty}^{+\infty} g(s-x)f(x)dx \right) ds$$

Per modo che la funzione

$$h(z) = \int_{-\infty}^{+\infty} g(z-x)f(x)dx$$

risulta essere la densità di probabilità della variabile aleatoria $\xi + \eta$.

Possiamo allora calcolare che

$$E(\zeta) = E(\xi + \eta) = \int_{-\infty}^{+\infty} s \left(\int_{-\infty}^{+\infty} g(s-x)f(x)dx \right) ds =$$
$$= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} sg(s-x)f(x)ds \right) dx =$$
$$= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} (x+t)g(t)f(x)dx \right) dt =$$
$$= \int_{-\infty}^{+\infty} g(t) \left(\int_{-\infty}^{+\infty} xf(x)dx \right) dt + \int_{-\infty}^{+\infty} f(x) \left(\int_{-\infty}^{+\infty} tg(t)dt \right) dx =$$
$$= E(\xi) \int_{-\infty}^{+\infty} g(t)dt + E(\eta) \int_{-\infty}^{+\infty} f(x)dx = E(\xi) + E(\eta)$$

1.7 *Prodotto di variabili aleatorie*

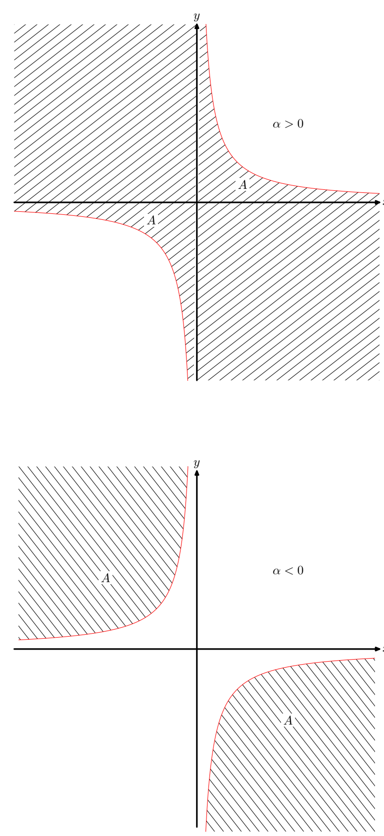


Figure 1.4:

Siano ξ e η due variabili aleatorie indipendenti le cui PDF sono f e g, rispettivamente, e sia

$$\zeta = \xi\eta$$

Avremo che

$$\mathcal{P}\left(\xi\eta \leq \alpha\right) = \int_A f(x)g(y)dxdy$$

dove

$$A = \{(x,y) \in \mathbb{R}^2 \ : \ xy \leq \alpha\}$$

è l'insieme tratteggiato nella figura 1.4;

Pertanto

$$\mathcal{P}\left(\xi\eta \leq \alpha\right) = \int_A f(x)g(y)dxdy =$$
$$= \int_{-\infty}^{0} \int_{\frac{\alpha}{x}}^{+\infty} f(x)g(y)dydx + \int_{0}^{+\infty} \int_{-\infty}^{\frac{\alpha}{x}} f(x)g(y)dydx =$$

posto $y = \frac{s}{x}$ da cui $dy = \frac{ds}{x}$

$$= \int_{-\infty}^{0}\int_{\alpha}^{-\infty} f(x)g\left(\frac{s}{x}\right)\frac{ds}{x}dx + \int_{0}^{+\infty}\int_{-\infty}^{\alpha} f(x)g\left(\frac{s}{x}\right)\frac{ds}{x}dx =$$
$$= \int_{-\infty}^{0}\left(-\int_{-\infty}^{\alpha} f(x)g\left(\frac{s}{x}\right)\frac{ds}{x}\right)dx + \int_{0}^{+\infty}\left(\int_{-\infty}^{\alpha} f(x)g\left(\frac{s}{x}\right)\frac{ds}{x}\right)dx =$$
$$\int_{-\infty}^{\alpha}\left(\int_{-\infty}^{+\infty} f(x)g\left(\frac{s}{x}\right)\frac{dx}{|x|}\right)ds$$

dal che si deduce che la PDF di ζ è data da

$$\varphi(s) = \int_{-\infty}^{+\infty} f(x)g\left(\frac{s}{x}\right)\frac{dx}{|x|}$$

Possiamo inoltre calcolare media e varianza di ζ come segue.

$$\mu_\zeta = \int_{-\infty}^{+\infty} s\left(\int_{-\infty}^{+\infty} f(x)g\left(\frac{s}{x}\right)\frac{dx}{|x|}\right)ds =$$
$$= \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} f(x)g\left(\frac{s}{x}\right)\frac{s}{|x|}ds\right)dx =$$
$$= \int_{0}^{+\infty}\left(\int_{-\infty}^{+\infty} f(x)g\left(\frac{s}{x}\right)\frac{s}{x}ds\right)dx - \int_{-\infty}^{0}\left(\int_{-\infty}^{+\infty} f(x)g\left(\frac{s}{x}\right)\frac{s}{x}ds\right)dx =$$

posto $t = \frac{s}{x}$ da cui $dt = \frac{ds}{x}$

$$= \int_{0}^{+\infty}\left(\int_{-\infty}^{+\infty} f(x)g(t)\,txdt\right)dx + \int_{-\infty}^{0}\left(\int_{-\infty}^{+\infty} f(x)g(t)\,txdt\right)dx =$$
$$= \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} f(x)g(t)\,txdt\right)dx = \mu_\xi\mu_\eta$$

Per quanto concerne la varianza avremo che

$$\sigma_\zeta^2 = \int_{-\infty}^{+\infty} s^2\left(\int_{-\infty}^{+\infty} f(x)g\left(\frac{s}{x}\right)\frac{dx}{|x|}\right)ds - \mu_\zeta^2$$

e

$$\int_{-\infty}^{+\infty} s^2\left(\int_{-\infty}^{+\infty} f(x)g\left(\frac{s}{x}\right)\frac{dx}{|x|}\right)ds = \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} f(x)g\left(\frac{s}{x}\right)\frac{s^2}{|x|}ds\right)dx =$$
$$= \int_{0}^{+\infty}\left(\int_{-\infty}^{+\infty} f(x)g\left(\frac{s}{x}\right)\frac{s^2}{x}ds\right)dx - \int_{-\infty}^{0}\left(\int_{-\infty}^{+\infty} f(x)g\left(\frac{s}{x}\right)\frac{s^2}{x}ds\right)dx =$$

posto $t = \frac{s}{x}$ da cui $dt = \frac{ds}{x}$

$$= \int_{0}^{+\infty}\left(\int_{-\infty}^{+\infty} f(x)g(t)\,\frac{x^2t^2}{x}xdt\right)dx + \int_{-\infty}^{0}\left(\int_{-\infty}^{+\infty} f(x)g(t)\,\frac{x^2t^2}{x}xdt\right)dx =$$
$$= \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} f(x)g(t)\,x^2t^2dt\right)dx =$$
$$= (\sigma_\xi^2 + \mu_\xi^2)(\sigma_\eta^2 + \mu_\eta^2)$$

Possiamo allora concludere che

$$\sigma_\zeta^2 + \mu_\zeta^2 = (\sigma_\xi^2 + \mu_\xi^2)(\sigma_\eta^2 + \mu_\eta^2)$$

1.7.1 Un caso particolare

Se ξ è una variabile aleatoria e $\alpha \in \mathbb{R}_+$ allora

$$\mathcal{P}(\alpha\xi \leq x) = \mathcal{P}\left(\xi \leq \frac{x}{\alpha}\right) = \int_{-\infty}^{\frac{x}{\alpha}} f(t)dt = \int_{-\infty}^{x} \frac{1}{\alpha} f\left(\frac{s}{\alpha}\right) ds$$

D'altro canto se $\alpha \in \mathbb{R}_-$

$$\mathcal{P}(\alpha\xi \leq x) = \mathcal{P}\left(\xi \geq \frac{x}{\alpha}\right) = \int_{\frac{x}{\alpha}}^{+\infty} f(t)dt = -\int_{-\infty}^{x} \frac{1}{\alpha} f\left(\frac{s}{\alpha}\right) ds$$

per cui la PDF di $\alpha\xi$ è data da

$$g(s) = \frac{1}{|\alpha|} f\left(\frac{s}{\alpha}\right)$$

1.8 Quoziente di variabili aleatorie

Siano η e ξ due variabili aleatorie positive e indipendenti le cui PDF sono g ed f, rispettivamente, e sia

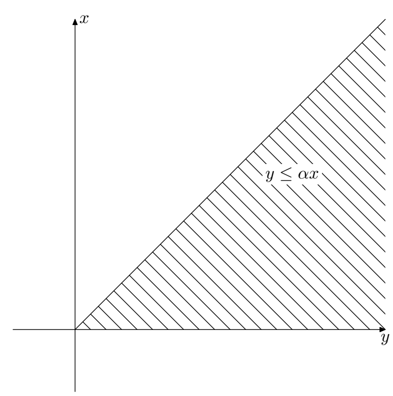
$$\zeta = \frac{\eta}{\xi}$$

Avremo che

$$\mathcal{P}\left(\frac{\eta}{\xi} \leq \alpha\right) = \mathcal{P}(\eta \leq \alpha\xi) = \int_A f(x)g(y)dxdy$$

dove

$$A = \{(x,y) \in \mathbb{R}^2 \ : \ y \leq \alpha x\}$$



Pertanto

$$\mathcal{P}\left(\frac{\eta}{\xi} \leq \alpha\right) = \int_A f(x)g(y)dxdy = \int_0^{+\infty} \int_0^{\alpha x} f(x)g(y)dydx =$$

posto $y = tx$ da cui $dy = xdt$

$$= \int_0^{+\infty} \int_0^{\alpha} f(x)g(tx)xdtdx = \int_0^{\alpha} \left(\int_0^{+\infty} xf(x)g(tx)dx \right) dt$$

dal che si deduce che la PDF di ζ è data da

$$\varphi(t) = \int_0^{+\infty} xf(x)g(tx)dx$$

1.8.1 Un caso particolare

Se ξ è una variabile aleatoria e $n \in \mathbb{R}_+$ allora

$$\mathcal{P}(\frac{\xi}{n} \leq \alpha) = \mathcal{P}(\xi \leq \alpha n) = \int_{-\infty}^{\alpha n} f(x)dx = \int_{-\infty}^{\alpha} nf(nt)dt$$

per cui la PDF di $\frac{\xi}{n}$ è data da

$$g(t) = nf(nt)$$

1.9 Distribuzioni di probabilità doppie

Siano $(\mathcal{U}_1, \mathcal{F}_1, \mathcal{P}_1)$ e $(\mathcal{U}_2, \mathcal{F}_2, \mathcal{P}_2)$ due spazi di probabilità consideriamo la variabile aleatoria che indichiamo con (ξ, η) definita sullo spazio $\mathcal{U}_1 \times \mathcal{U}_2$ mediante la

$$F(x, y) = \mathcal{P}(\xi \leq x, \eta \leq y) = \int_{-\infty}^{x} \left(\int_{-\infty}^{y} f(t, s) ds \right) dt$$

F è la distribuzione cumulativa di probabilità della variabile (ξ, η) ed f è la sua funzione distribuzione di probabilità

Se f è continua possiamo affermare che

$$\frac{\partial^2 F}{\partial x \partial y} = f(x, y)$$

Naturalmente devono essere verificate le seguenti condizioni:

-
$$f(x, y) \geq 0$$

-
$$\int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(t, s) ds \right) dt = 1$$

Inoltre se

$$F_1(x) = \mathcal{P}(\xi \leq x) = \int_{-\infty}^{x} \left(\int_{-\infty}^{+\infty} f(t, s) ds \right) dt$$
$$F_2(y) = \mathcal{P}(\eta \leq y) = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{y} f(t, s) ds \right) dt = \int_{-\infty}^{y} \left(\int_{-\infty}^{+\infty} f(t, s) dt \right) ds$$

F_1 ed F_2 sono le distribuzioni cumulative delle variabili aleatorie ξ e η, rispettivamente le cui funzioni di distribuzione sono date da

$$\varphi(t) = \int_{-\infty}^{+\infty} f(t, s) ds$$
$$\psi(s) = \int_{-\infty}^{+\infty} f(t, s) dt$$

Nel caso in cui le variabili aleatorie ξ e η siano indipendenti, allora (ξ, η) ha una distribuzione di probabilità

$$f(t, s) = \varphi(t)\psi(s)$$

dove φ e ψ sono le funzioni di distribuzione di ξ e η, rispettivamente.

È utile ricordare che la probabilità della variabile aleatoria ξ condizionata alla variabile aleatoria η si può definire mediante la

$$\mathcal{P}(\xi \leq x | \eta \leq y) = \int_{-\infty}^{x} \left(\int_{-\infty}^{y} \frac{f(t, s)}{\psi(s)} ds \right) dt$$

per cui $\frac{f(t,s)}{\psi(s)}$ è la sua funzione di distribuzione di probabilità.

Possiamo giustificare la definizione osservando che:

$$
\begin{aligned}
\mathcal{P}(\xi \leq x, y \leq \eta \leq y+k) = \frac{\int_{-\infty}^{x}\int_{y}^{y+k} f(t,s)dsdt}{\int_{-\infty}^{+\infty}\int_{y}^{y+k} f(t,s)dsdt} = \\
= \frac{\int_{-\infty}^{x}\int_{y}^{y+k} f(t,s)dsdt}{\int_{y}^{y+k}\int_{-\infty}^{+\infty} f(t,s)dtds} = \frac{\int_{-\infty}^{x}\int_{y}^{y+k} f(t,s)dsdt}{\int_{y}^{y+k}\psi(s)ds} = \\
\approx \frac{\int_{-\infty}^{x} f(t,y)kdt}{\psi(y)k} = \int_{-\infty}^{x}\frac{f(t,y)k}{\psi(y)k}dt = \int_{-\infty}^{x}\frac{f(t,y)}{\psi(y)}dt
\end{aligned}
$$

1.10 *Normalizzazione di una variabile aleatoria.*

Sia ξ una variabile aleatoria di media μ e di varianza σ^2 con distribuzione di probabilità φ.

e consideriamo la variabile aleatoria

$$\xi^* = \frac{\xi - \mu}{\sigma}$$

Per le proprietà di media e varianza possiamo affermare che ξ^* è una variabile normalizzata (o standardizzata), intendendo con ciò che ξ^* ha media 0 e varianza 1.

Allo scopo di determinare la funzione di distribuzione di ξ^* osserviamo che

$$
\begin{aligned}
&\mathcal{P}\left(a \leq \xi^* \leq b\right) = \\
&= \mathcal{P}\left(a \leq \frac{\xi-\mu}{\sigma} \leq b\right) = \mathcal{P}(\mu + \sigma a \leq \xi \leq \mu + \sigma b) = \int_{\mu+\sigma a}^{\mu+\sigma b} \varphi(s)ds = \\
&= \int_{a}^{b} \sigma\varphi\left(\mu + \sigma t\right)\right) dt
\end{aligned}
$$

Pertanto la variabile aleatoria

$$\xi^* = \frac{\xi - \mu}{\sigma}$$

ha una PDF definita da

$$\psi(t) = \sigma\varphi\left(\mu + \sigma t\right)\right)$$

e possiamo allora verificare che

$$
\begin{aligned}
\int_{-\infty}^{+\infty} \psi(t)dt &= \int_{-\infty}^{+\infty} \sigma\varphi\left(\mu+\sigma t\right)dt = \int_{-\infty}^{+\infty} \varphi(s)ds = 1 \\
\int_{-\infty}^{+\infty} t\psi(t)dt &= \int_{-\infty}^{+\infty} t\sigma\varphi\left(\mu+\sigma t\right)dt = \int_{-\infty}^{+\infty} \left(\frac{s-\mu}{\sigma}\right)\varphi(s)ds = \\
&= \frac{1}{\sigma}\left(\int_{-\infty}^{+\infty} s\varphi(s)ds - \mu\int_{-\infty}^{+\infty} \varphi(s)ds\right) = 0 \\
\int_{-\infty}^{+\infty} t^2\psi(t)dt &= \int_{-\infty}^{+\infty} t^2\sigma\varphi\left(\mu+\sigma t\right)dt = \int_{-\infty}^{+\infty} \left(\frac{s-\mu}{\sigma}\right)^2\varphi(s)ds = \\
&= \frac{1}{\sigma^2}\int_{-\infty}^{+\infty} (s-\mu)^2\varphi(s)ds = 1
\end{aligned}
$$

2. Qualche Distribuzione di Probabilità

Le funzioni di distribuzione di probabilità sono fondamentali per descrivere il comportamento delle variabili aleatorie che ci interessano.

Ogni variabile aleatoria ha una sua distribuzione e per definirne la proprietà è utile fare riferimento ad alcune distribuzioni note che sono in grado di descrivere la maggior parte delle variabili aleatorie con cui normalmente si lavora.

2.1 La distribuzione uniforme

La più semplice funzione di distribuzione di probabilità è quella di una variabile aleatoria che restituisce un valore scelto in un intervallo $[a, b]$ con il criterio di equiprobabilità.

Abbiamo già visto che in tal caso

$$\mathcal{P}(x \leq \xi \leq x + h) = \frac{h}{b - a}$$

e che la sua distribuzione di densità è

$$\boxed{\varphi(t) = \begin{cases} \frac{1}{b-a} & t \in [a, b] \\ 0 & \text{altrove} \end{cases}}$$

La funzione generatrice dei momenti si calcola mediante la

$$M_\xi(t) = \frac{1}{b - a} \int_a^b e^{tx} dx = \frac{e^{tb} - e^{ta}}{t(b - a)}$$

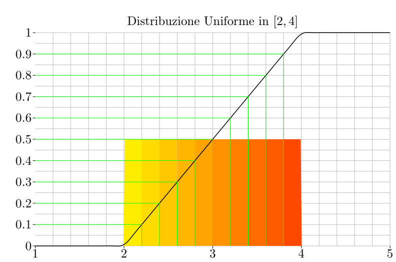


Figure 2.1: PDF e CDF di una variabile aleatoria Uniforme

Se ne ricava subito che

$$\mu = \frac{b + a}{2}$$
$$\sigma^2 = \frac{(b - a)^2}{12}$$

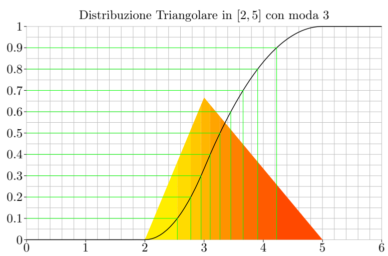


Figure 2.2: PDF e CDF di una variabile aleatoria Triangolare

2.2 *La distribuzione triangolare*

La distribuzione triangolare è utile per definire una variabile aleatoria che assuma valori compresi tra a e b ed abbia una moda c. La funzione distribuzione di probabilità triangolare si definisce mediante la

$$\boxed{\phi(t) = \begin{cases} 0 & t < a \\ \frac{2(t-a)}{(b-a)(c-a)} & a \leq t < c \\ \frac{2(b-t)}{(b-a)(b-c)} & c < t \leq b \\ 0 & t > b \end{cases}}$$

Si calcola facilmente che il valor medio è

$$\mu = \frac{a+b+c}{3}$$

mentre la varianza è data da

$$\sigma^2 = \frac{a^2+b^2+c^2-bc-ab-ac}{18}$$

e la funzione generatrice dei momenti è

$$M_\xi(t) = \frac{e^{ta}(b-c) - e^{tc}(b-a) + e^{tb}(c-a)}{t^2(b-a)(c-a)(b-c)}$$

2.3 *Alcune importanti distribuzioni discrete*

2.3.1 *La distribuzione binomiale di Bernoulli*

Definizione 2.1 *Chiamiamo prova bernoulliana un esperimento che ha due soli possibili esiti:*

- *Successo, cui associamo il valore* 1 *con probabilità* p
- *Insuccesso, cui associamo il valore* 0 *con probabilità* q

essendo ovviamente $p + q = 1$.

Chiamiamo variabile aleatoria bernoulliana la variabile aleatoria ξ *che restituisce il numero di successi che si sono verificati su* n *prove ripetute (lanci) dell'esperimento.*

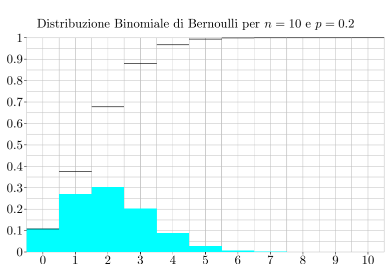


Figure 2.3: PDF e CDF di una variabile aleatoria Binomiale (di Bernoulli).

Possiamo calcolare la probabilità che la variabile aleatoria ξ assuma il valore k mediante la

$$\mathcal{P}(\xi = k) = \binom{n}{k} p^k q^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}$$

Per giustificare la formula precedente descriviamo la successione di n prove ripetute con una stringa di elementi che assumono il valore

1 oppure 0 a seconda che la corrispondente prova abbia avuto o no successo.

$$\boxed{0}\boxed{1}\boxed{1}\boxed{0}\boxed{0}\boxed{1}\boxed{0}\boxed{0}\boxed{1}$$

oppure

$$\boxed{0}\boxed{0}\boxed{1}\boxed{0}\boxed{1}\boxed{0}\boxed{0}\boxed{0}\boxed{1}\boxed{0}\boxed{1}$$

affinchè ci siano k successi la stringa dovrà contenere esattamente k volte il valore 1 (ed $n-k$ volte il valore 0) e quindi, poichè in ogni elemento 1 si presenta con probabilità p mentre il valore 0 compare con probabilità q, una stringa con k successi avrà una probabilità di comparire uguale a

$$p^k q^{n-k}$$

d'altro canto, poichè siamo unicamente interessati a contare il numero di successi, e non l'ordine con cui si verificano, dovremo tener conto che si possono ottenere, ad esempio, k successi su n prove in tanti modi diversi

$$\boxed{0}\boxed{1}\boxed{0}\boxed{0}\boxed{0}\boxed{1}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}$$

$$\boxed{0}\boxed{0}\boxed{1}\boxed{1}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}$$

il cui numero è dato dalle combinazioni di n oggetti a k a k e cioè

$$\binom{n}{k}$$

(Ciascuna combinazione è individuata dalla sequenza dei k numeri, compresi tra 1 ed n, che indicano la posizione dei successi.)

Possiamo calcolare la media della variabile bernoulliana ξ osservando che la media in ciascuna prova è

$$1 \cdot p + 0 \cdot q = p$$

e su n esperimenti, essendo la media lineare, avremo

$$\mu = E(\xi) = np$$

La varianza della variabile bernoulliana ξ in ciascuna prova è

$$(1-p)^2 \cdot p + (0-p)^2 q = q^2 p + p^2 q = pq(p+q) = pq$$

e su n esperimenti per le proprietà della varianza avremo

$$\sigma^2 = E((\xi - \mu)^2) = npq$$

e

$$\sigma = \sqrt{npq}$$

Alternativamente possiamo calcolare la media e la varianza di una variabile aleatoria Bernoulliana ξ usando direttamente la definizione:

$$
\begin{aligned}
\mu = E(\xi) = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} = \sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k} = \\
= \sum_{k=1}^{n} k \frac{n(n-1)\cdots(n-(k-1))}{k!} p^k q^{n-k} = \\
= np \sum_{k=1}^{n} \frac{(n-1)(n-2)\cdots(n-(k-1))}{(k-1)!} p^{k-1} q^{n-1-(k-1)} = \\
= np \sum_{k=0}^{n-1} \frac{(n-1)(n-2)\cdots(n-1-(k-1))}{k!} p^k q^{n-1-k} = \\
= np(p+q)^{n-1} = np
\end{aligned}
$$

$$
\begin{aligned}
\sigma^2 = E((\xi-\mu)^2) = \sum_{k=0}^{n} (k-np)^2 \binom{n}{k} p^k q^{n-k} = \\
= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k q^{n-k} - (np)^2 = \sum_{k=1}^{n} k^2 \binom{n}{k} p^k q^{n-k} - (np)^2 = \\
= np \sum_{k=1}^{n} k \frac{(n-1)(n-2)\cdots(n-(k-1))}{(k-1)!} p^{k-1} q^{n-1-(k-1)} - (np)^2 = \\
= np \sum_{k=0}^{n-1} (k+1) \frac{(n-1)(n-2)\cdots(n-1-(k-1))}{k!} p^k q^{n-1-k} - (np)^2 = \\
= np \left(\sum_{k=0}^{n-1} k \binom{n-1}{k} p^k q^{n-1-k} + \sum_{k=0}^{n-1} \binom{n-1}{k} p^k q^{n-1-k} \right) - (np)^2 = \\
= np((n-1)p+1) - (np)^2 = np(np+q) - (np)^2 = npq
\end{aligned}
$$

Per calcolare la funzione generatrice dei momenti possiamo procedere come segue

$$
\begin{aligned}
M_\xi(t) = E(e^{t\xi}) = \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k q^{n-k} = \\
= \sum_{k=0}^{n} \binom{n}{k} \left(pe^t\right)^k q^{n-k} = (pe^t+q)^n
\end{aligned}
$$

la funzione densità di probabilità (Probability Density Function , PDF) di una variabile aleatoria di Bernoulli ξ è definita da

$$
P(\xi = k) = \binom{n}{k} p^k q^{n-k}
$$

mentre la funzione di distribuzione cumulativa (Cumulative Distribution Function, CDF) è

$$
F(\nu) = \sum_{k=0}^{\nu} P(\xi = k) = \sum_{k=0}^{\nu} \binom{n}{k} p^k q^{n-k}
$$

Ad esempio si ha

$$F(0) = \sum_{k=0}^{0} P(\xi = k) = P(\xi = 0) = q^n$$

ed anche

$$F(n) = \sum_{k=0}^{n} P(\xi = k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1$$

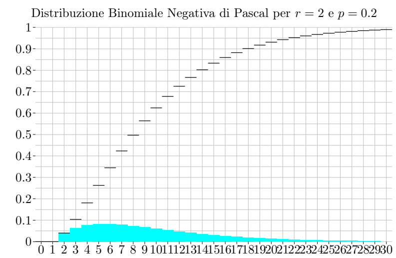


Figure 2.4: PDF e CDF di una variabile aleatoria Binomiale Negativa (di Pascal).

2.3.2 *La distribuzione binomiale negativa di Pascal*

Consideriamo un esperimento bernoulliano, consideriamo cioè una serie di prove ripetute con due soli possibili esiti: successo con probabilità p ed insuccesso con probabilità q.

Consideriamo la variabile aleatoria che restituisce il minimo numero ξ di tentativi necessari per ottenere r successi.

Possiamo allora vedere che la probabilità $\mathcal{P}(\xi = k)$ che si ottengano r successi al tentativo k si può calcolare considerando che

- al tentativo k si è verificato un successo (che ha probabilità p)
- nelle precedenti $k-1$ prove si sono verificati $r-1$ successi e $k-1-(r-1)$ insuccessi (con probabilità $\binom{k-1}{r-1} p^{r-1} q^{k-r}$)

Pertanto

$$\mathcal{P}(\xi = k) = p \binom{k-1}{r-1} p^{r-1} q^{k-r} = \binom{k-1}{r-1} p^r q^{k-r}$$

definisce la funzione densità di probabilità della distribuzione di Pascal.

Talvolta si considera, in luogo di ξ, la variabile η che restituisce il numero di fallimenti che precedono il successo r-esimo. In tal caso si ha $h = k - r$ e

$$\mathcal{P}(\eta = y) = \binom{h+r-1}{r-1} p^r q^h$$

Possiamo calcolare che

$$\mu_\xi = \frac{r}{p} \quad , \quad \sigma^2_\xi = \frac{rq}{p^2} \quad , \quad M_\xi(t) = \frac{pe^t}{(1-qe^t)^r}$$

e

$$\mu_\eta = \frac{r(1-p)}{p} \quad , \quad \sigma^2_\eta = \frac{r(1-p)}{p^2}$$

A titolo di esempio vediamo come è possibile calcolare la media μ_ξ e la varianza σ_ξ;

$$
\begin{aligned}
\mu_\xi = \sum_{k=r}^{+\infty} kp\binom{k-1}{r-1}p^{r-1}q^{k-r} &= \sum_{k=r}^{+\infty} p^r \frac{k(k-1)\dots(k-r+1)}{(r-1)!}q^{k-r} = \\
&= \frac{p^r}{(r-1)!}\sum_{k=r}^{+\infty} k(k-1)\dots(k-r+1)q^{k-r} = \\
&= \frac{p^r}{(r-1)!}\sum_{k=r}^{+\infty} \frac{d^r}{dq^r}q^k = \frac{p^r}{(r-1)!}\frac{d^r}{dq^r}\sum_{k=r}^{+\infty} q^k = \\
= \frac{p^r}{(r-1)!}\frac{d^r}{dq^r}\sum_{k=0}^{+\infty} q^k &= \frac{p^r}{(r-1)!}\frac{d^r}{dq^r}\frac{1}{1-q} = \frac{p^r}{(r-1)!}r!\frac{1}{(1-q)^{r+1}} = \\
&= \frac{p^r}{(r-1)!}\frac{r!}{p^{r+1}} = \frac{r}{p}
\end{aligned}
$$

mentre

$$
\begin{aligned}
\sum_{k=r}^{+\infty} k^2 p\binom{k-1}{r-1}p^{r-1}q^{k-r} &= \sum_{k=r}^{+\infty} kp^r \frac{k(k-1)\dots(k-r+1)}{(r-1)!}q^{k-r} = \\
= \frac{p^r}{(r-1)!}\sum_{k=r}^{+\infty} kk(k-1)\dots(k-r+1)q^{k-r} &= \frac{p^r}{(r-1)!}\sum_{k=r}^{+\infty} k\frac{d^r}{dq^r}q^k = \\
&= \frac{p^r}{(r-1)!}\frac{d^r}{dq^r}\sum_{k=r}^{+\infty} kq^k = \frac{p^r}{(r-1)!}\frac{d^r}{dq^r}\sum_{k=0}^{+\infty} kq^k = \\
&= \frac{p^r}{(r-1)!}\frac{d^r}{dq^r}\frac{q}{(1-q)^2} = \frac{p^r}{(r-1)!}\left(\frac{1}{(1-q)^2} - \frac{1}{(1-q)}\right) = \\
&= \frac{p^r}{(r-1)!}\left(\frac{(r+1)!}{p^{r+2}} - \frac{r!}{p^{r+1}}\right) = \\
&= \frac{p^r}{(r-1)!}r!\frac{(r+q)}{p^{r+2}} = \frac{r(r+q)}{p^2}
\end{aligned}
$$

da cui

$$
\sigma_\xi = \sum_{k=r}^{+\infty} k^2 p\binom{k-1}{r-1}p^{r-1}q^{k-r} - \mu_\xi^2 = \frac{r(r+q)}{p^2} - \frac{r^2}{p^2} = \frac{rq}{p^2}
$$

2.3.3 *La distribuzione geometrica*

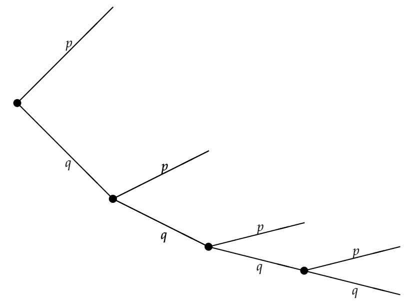
Consideriamo una prova con probabilità di successo p ; ripetiamola indefinitamente sotto l'ipotesi che

- p rimane costante:
- l'esito della prova non dipende dalle prove precedentemente effettuate.

Ad esempio possiamo considerare un tiratore che ha la capacità di colpire il bersaglio con probabilità p ad ogni tiro o una lampada che può guastarsi con probabilità p ad ogni accensione.

Sia ξ la variabile aleatoria che restituisce il numero del primo tentativo in cui la prova ha successo. Avremo che

$$\mathcal{P}(\xi = k) = (1-p)^{k-1} p$$



in quanto la prova ha avuto esito negativo (la probabilità di insuccesso è $1-p$) per $k-1$ volte ed ha avuto successo la k-esima volta.

Pertanto la distribuzione di probabilità della variabile aleatoria ξ è definita da

$$\boxed{\varphi(k) = \mathcal{P}(\xi = k) = (1-p)^{k-1} p}$$

e si verifica subito che

$$\sum_{1}^{+\infty} p(1-p)^{k-1} = p \sum_{0}^{+\infty} (1-p)^k = p \frac{1}{1-(1-p)} = 1$$

Inoltre

$$\mu = \sum_{1}^{+\infty} kp(1-p)^{k-1} = p \sum_{1}^{+\infty} k(1-p)^{k-1} =$$
$$= p \sum_{1}^{+\infty} -\frac{d}{dp}(1-p)^k = -p\frac{d}{dp} \sum_{1}^{+\infty} (1-p)^k = -p\left(-\frac{1}{p^2}\right) = \frac{1}{p}$$

mentre

$$\sum_{1}^{+\infty} k^2 p(1-p)^{k-1} = \sum_{1}^{+\infty} (k^2+k)p(1-p)^{k-1} - \sum_{1}^{+\infty} kp(1-p)^{k-1} =$$
$$= p \sum_{1}^{+\infty} \frac{d^2}{dp^2}(1-p)^{k+1} - \mu = p\frac{d^2}{dp^2} \sum_{1}^{+\infty} (1-p)^{k+1} - \mu =$$
$$= p\frac{d^2}{dp^2}\left(\frac{1}{p} - 2 - p\right) - \mu = \frac{2}{p^2} - \mu$$

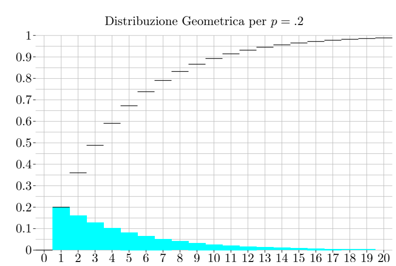


Figure 2.5: PDF e CDF di una variabile aleatoria Geometrica

per cui

$$\sigma^2 = \sum_{1}^{+\infty} k^2 p(1-p)^{k-1} - \mu^2 = \frac{2}{p^2} - \mu - \mu^2 = \frac{1-p}{p^2}$$

Per quanto riguarda la funzione generatrice dei momenti

$$M_\xi(t) = E(e^{t\xi}) = \sum_{k=1}^{+\infty} e^{tk} p(1-p)^{k-1} = \frac{p}{1-p} \sum_{k=1}^{+\infty} (e^t(1-p))^k =$$
$$= \frac{p}{1-p}\left(\frac{e^t(1-p)}{1-e^t(1-p)}\right) = \frac{pe^t}{1-e^t(1-p)}$$

2.3.4 *La distribuzione di Poisson*

Consideriamo un centralino telefonico che in media riceve λ chiamate all'ora e supponiamo di voler determinare la probabilità che riceva k chiamate in un'ora.

Suddividiamo l'ora in n parti uguali ciascuna della durata di $\frac{1}{n}$; durante ciascuno degli n periodi di durata $\frac{1}{n}$ la probabilità che si riceva una chiamata è λ/n, pertanto la probabilità che si ricevano k chiamate si può ottenere considerando la probabilità che una variabile aleatoria binomiale relativa ad n prove ripetute con probabilità di successo $\frac{\lambda}{n}$ assuma valore k Sia quindi ξ la variabile aleatoria che restituisce il numero di successi ottenuti.

Avremo che

$$\begin{aligned}\varphi(k) = \mathcal{P}(\xi = k) = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} &= \\ = \lambda^k\binom{n}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n}\left(1-\frac{\lambda}{n}\right)^{-k} &=\end{aligned}$$

Se ora consideriamo di far tendere n a $+\infty$ avremo

$$\left(1-\frac{\lambda}{n}\right)^{n} \to e^{-\lambda}$$

$$\left(1-\frac{\lambda}{n}\right)^{-k} \to 1$$

$$\begin{aligned}\binom{n}{k}\left(\frac{1}{n}\right)^k = \frac{n!}{k!(n-k)!}\frac{1}{n^k} = \frac{1}{k!}\frac{n!}{(n-k)!n^k} &= \\ \frac{1}{k!}\frac{n^n e^{-n}\sqrt{2\pi n}}{(n-k)^{n-k}e^{-(n-k)}\sqrt{2\pi(n-k)}} &= \\ = \frac{1}{k!}\frac{1}{e^k}\sqrt{\frac{n}{n-k}}\frac{n^n}{(n-k)^n} = \frac{1}{k!}\frac{1}{e^k}\frac{n^n}{(n-k)^n}\sqrt{\frac{n}{n-k}} &= \\ \frac{1}{k!}\frac{1}{e^k}\left(1-\frac{k}{n}\right)^{-n}\sqrt{\frac{n}{n-k}} \to \frac{1}{k!}\end{aligned}$$

non appena si tenga conto che

$$\left(1-\frac{k}{n}\right)^{-n} \to e^k \qquad \text{e} \qquad \sqrt{\frac{n}{n-k}} \to 1$$

Pertanto la funzione distribuzione di probabilità della variabile aleatoria considerata è data da

$$\boxed{\varphi(k) = \frac{1}{k!}\lambda^k e^{-\lambda}}$$

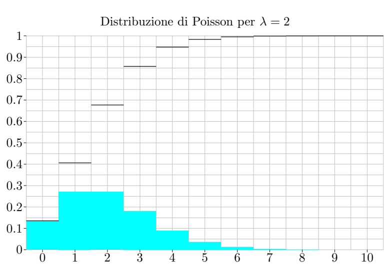


Figure 2.6: PDF e CDF di una variabile aleatoria di Poisson.

In generale chiamiamo variabile aleatoria di Poisson la variabile che restituisce il numero di eventi accaduti nell'unità di tempo, noto il fatto che il numero medio di eventi che accadono nell'unità di tempo è λ.

Per calcolare media, varianza ed i momenti della distribuzione di Poisson è utile calcolare la funzione generatrice dei momenti.

$$M_\xi(t) = \sum_{k=0}^{+\infty} e^{tk} \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=0}^{+\infty} \frac{(\lambda e^t)^k}{k!} e^{-\lambda} = e^{-\lambda} e^{\lambda(e^t-1)}$$

e le sue derivate

$$\frac{d}{dt} M_\xi(t) = e^{\lambda(e^t-1)} \lambda e^t$$
$$\frac{d^2}{dt^2} M_\xi(t) = e^{\lambda(e^t-1)} \lambda^2 e^t + e^{\lambda(e^t-1)} \lambda e^t$$

Calcolando in $t = 0$ si ottiene

$$\frac{d}{dt} M_\xi(0) = \lambda = \mu_1' = \mu$$
$$\frac{d^2}{dt^2} M_\xi(0) = \lambda^2 + \lambda = \mu_2'$$

da cui

$$\mu = \mu_1' = \lambda$$
$$\sigma^2 = \mu_2' - \mu^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

2.3.5 *Somma di variabili poissoniane*

Sui può verificare che la somma di due variabili aleatorie di Poisson di media λ e μ è ancora una variabile di Poisson di media $\lambda + \mu$, infatti se

$$\varphi(k) = \frac{1}{k!} \lambda^k e^{-\lambda} \qquad \text{e} \qquad \psi(k) = \frac{1}{k!} \mu^k e^{-\mu}$$

possiamo calcolare la densità di probabilità della variabile somma mediante la

$$\theta(k) = \sum_{h=0}^{n} \varphi(h)\psi(k-h) = \sum_{h=0}^{n} \frac{1}{h!} \lambda^h e^{-\lambda} \frac{1}{(k-h)!} \mu^{k-h} e^{-\mu} =$$
$$= e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{h=0}^{n} \frac{k!}{h!(k-h)!} \lambda^h \mu^{k-h} =$$
$$= e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{h=0}^{n} \binom{k}{h} \lambda^h \mu^{k-h} = e^{-(\lambda+\mu)} \frac{1}{k!} (\lambda + \mu)^k$$

2.3.6 *La distribuzione multinomiale*

Consideriamo un esperimento che possa avere k possibili esiti, che indichiamo con

$$A_1, A_2, ..., A_k$$

con probabilità

$$p_1, p_2, ..., p_k \quad , \quad p_1 + p_2 + ... + p_k = 1$$

e supponiamo di replicarlo per n volte; consideriamo la variabile aleatoria ξ che restituisce la $n-$pla di valori

$$n_1, n_2, ..., n_k \quad , \quad n_1 + n_2 + ... + n_k = n$$

dove n_i è il numero di volte in cui si è verificato l'evento A_i.

La funzione distribuzione di probabilità di ξ è data da

$$\boxed{\begin{aligned}\varphi(n_1, n_2, ..., n_k) &= \mathcal{P}(\xi_1 = n_1, \xi_2 = n_2, ..., \xi_k = n_k) = \\ &= \frac{n!}{n_1!n_2!......n_k!} p_1^{n_1} p_2^{n_2} ......p_k^{n_k}\end{aligned}}$$

Infatti su n tentativi si sono verificati n_1 successi con probabilità p_1, sui restanti $n - n_1$ tentativi si sono verificati n_2 successi con probabilià p_2 e così via fino at ottenere n_k successi su $(n - n_i - n_2 - \cdots n_{k-1}$ tentativi per cui

$$\begin{aligned}\varphi(n_1, n_2, ..., n_k) &= \mathcal{P}(\xi_1 = n_1, \xi_2 = n_2, ..., \xi_k = n_k) = \\ &= \frac{n!}{n_1!(n-n_1)!} p_1^{n_1} \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} p_2^{n_2} \cdots \\ &\cdots \frac{(n-n_1-n_2-n_{k-1})!}{(n_k)!(n-n_1-n_2-\cdots n_k)!} p_k^{n_k} = \\ &(\text{dal momento che } n - n_1 - n_2 - \cdots n_k = 0) \\ &= \frac{n!}{n_1!n_2!......n_k!} p_1^{n_1} p_2^{n_2} ......p_k^{n_k}\end{aligned}$$

2.3.7

Sia $\xi = (\xi_1, \xi_2, \ldots \xi_k)$ una variabile aleatoria distribuita multinomialmente relativa al caso in cui gli eventi $A_1, A_2, \ldots, A_k$ hanno probabilità di accadimento $p_1, p_2, \ldots p_k$ e sia $\eta = \eta_1 + \eta_2 +, \cdots, \eta_k$ una variabile aleatoria le cui componenti η_j sono variabili indipendenti con densità di Poisson di media $\lambda_k = np_k$ Allora

$$\begin{aligned}&\mathcal{P}(\xi_1 = n_1, \xi_2 = n_2, \ldots, \xi_k = n_k) = \\ &\qquad = \mathcal{P}(\eta_1 = n_1, \eta_2 = n_2, \ldots, \eta_n = n_k | \eta_1 + \eta_2 + \cdots + \eta_k = n)\end{aligned}$$

Infatti, dal momento che le η_j son indipendenti si ha che e che

$$p_1 + p_2 + \cdots p_k = 1$$

$$\mathcal{P}(\eta_1 = n_1, \eta_2 = n_2, \dots, \eta_k = n_k) =$$
$$\left(\frac{(np_1)^{n_1} e^{-np_1}}{n_1!}\right)\left(\frac{(np_2)^{n_2} e^{-np_2}}{n_2!}\right)\cdots\left(\frac{(np_k)^{n_k} e^{-np_k}}{n_k!}\right) =$$
$$= \left(\frac{(n)^{n_1+n_2+\cdots n_k} p_1^{n_1} p_2^{n_2}\cdots p_k^{n_k}}{n_1!\, n_2!\,\cdots\, n_k!}\right) e^{-n(p_1+p_2+p_k)} =$$
$$= \left(\frac{(n)^{n_1+n_2+\cdots n_k} p_1^{n_1} p_2^{n_2}\cdots p_k^{n_k}}{n_1!\, n_2!\,\cdots\, n_k!}\right) e^{-n} =$$

D'altra parte

$$\mathcal{P}(\eta_1 = n_1, \eta_2 = n_2, \dots, \eta_k = n_k | \eta_1 + \eta_2 + \cdots + \eta_k = n) =$$
$$= \frac{\left(\frac{(n)^n p_1^{n_1} p_2^{n_2}\cdots p_k^{n_k}}{n_1!\, n_2!\,\cdots\, n_k!}\right) e^{-n}}{\frac{n^n e^{-n}}{n!}} =$$
$$= \frac{n!}{n_1!\, n_2!\,\cdots\, n_k!} p_1^{n_1} p_2^{n_2}\cdots p_k^{n_k} =$$
$$= \mathcal{P}(\xi_1 = n_1, \xi_2 = n_2, \dots, \xi_k = n_k)$$

2.3.8 *La distribuzione ipergeometrica*

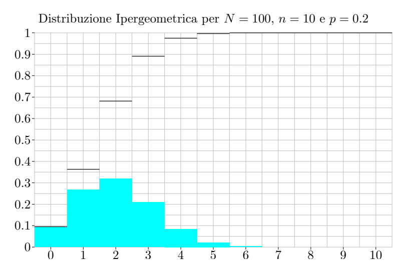


Figure 2.7: PDF e CDF di una variabile aleatoria Ipergeometrica.

Consideriamo un'urna contenente b palline nere e w palline bianche e supponiamo di estrarre per n volte una pallina rimettendola. dopo ogni estrazione, nell'urna.

Consideriamo la variabile aleatoria ξ che restituisce il numero di volte in cui si è estratta una pallina nera; allora la densità di probabilità di ξ si può calcolare mediante la

$$\varphi(k) = \mathcal{P}(\xi = k) = \binom{n}{k}\left(\frac{b}{b+w}\right)^k \left(\frac{w}{b+w}\right)^{n-k} = \binom{n}{k}\frac{b^k w^{n-k}}{b+w}$$

non appena si ricordi la distribuzione binomiale e si tenga presente che

$$p = \frac{b}{b+w} \quad , \quad q = \frac{w}{b+w}$$

Qualora l'esperimento si ripeta senza rimettere la pallina estratta nell'urna, (campionamento senza ripetizione), si può vedere che la densità di probabilità della nuova variabile aleatoria ξ che conta il numero delle palline nere estratte è

$$\boxed{\varphi(k) = \mathcal{P}(\xi = k) = \frac{\binom{b}{k}\binom{w}{n-k}}{\binom{b+w}{n}}}$$

Infatti il denominatore conta quante $n - ple$ di palline si possono formare avendo a disposizione $b + w$ palline, mentre a numeratore c'è il numero delle $n-$ple che contengono esattamente k palline nere che si possono ottenere conbinando una $k - pla$ di palline nere, in numero di $\binom{b}{k}$, con una $(n-k) - pla$ di palline bianche, in numero di $\binom{w}{n-k}$.

Si calcola anche che

$$\mu = \frac{nb}{b+w} \qquad , \qquad \sigma^2 = \frac{nbw(b+w-n)}{(b+w)^2(b+w-1)}$$

Infatti

Ricordiamo l'identità di Vandermonde:

$$\binom{m+n}{k} = \sum_{h=0}^{k} \binom{m}{h}\binom{n}{k-h}$$

infatti:

- $\binom{m+n}{k}$ numero dei modi con cui si possono scegliere k elementi tra $m+n$
- $\binom{m}{h}$ numero dei modi con cui si possono scegliere h elementi tra m
- $\binom{n}{k-h}$ numero dei modi con cui si possono scegliere $k-h$ elementi tra n

k elementi tra $m+n$ si scelgono prendendone h tra i primi m e $k-h$ tra gli altri n. quindi per h fissato ci sono $\binom{m}{h}\binom{n}{k-h}$ possibili scelte. Sommando su h si trovano tutte e si ottiene la formula.

$$\mu = \sum_{k=0}^{n} k\mathcal{P}(\xi = k) = \sum_{k=0}^{n} k\frac{\binom{b}{k}\binom{w}{n-k}}{\binom{b+w}{n}} =$$

$$= \frac{1}{\binom{b+w}{n}} \sum_{k=1}^{n} b\binom{b-1}{k-1}\binom{w}{n-k} = \frac{1}{\binom{b+w}{n}} \sum_{k=0}^{n-1} b\binom{b-1}{k}\binom{w}{n-k-1} =$$

$$= b\frac{1}{\binom{b+w}{n}}\binom{b+w-1}{n-1} = b\frac{n!(b+w-n)!}{(b+w)!}\frac{(b+w-1)!}{(n-1)!(b+w-n)!} =$$

$$= \frac{bn}{b+w}$$

Inoltre

$$\sum_{k=0}^{n} k^2\mathcal{P}(\xi = k) = \sum_{k=0}^{n} k^2\frac{\binom{b}{k}\binom{w}{n-k}}{\binom{b+w}{n}} =$$

$$= \frac{1}{\binom{b+w}{n}} \sum_{k=1}^{n} kb\binom{b-1}{k-1}\binom{w}{n-k} =$$

$$= \frac{1}{\binom{b+w}{n}} \sum_{k=0}^{n-1} (k+1)b\binom{b-1}{k}\binom{w}{n-k-1} =$$

$$= \frac{b}{\binom{b+w}{n}} \left(\sum_{k=1}^{n-1} k\binom{b-1}{k}\binom{w}{n-k-1} + \sum_{k=0}^{n-1}\binom{b-1}{k}\binom{w}{n-k-1}\right) =$$

$$= \frac{b}{\binom{b+w}{n}} \left(\sum_{k=1}^{n-1} (b-1)\binom{b-2}{k-1}\binom{w}{n-k-1} + \binom{b+w-1}{n-1}\right) =$$

$$= \frac{b}{\binom{b+w}{n}} \left(\sum_{k=0}^{n-2} (b-1)\binom{b-2}{k}\binom{w}{n-k-2} + \binom{b+w-1}{n-1}\right) =$$

$$= \frac{b}{\binom{b+w}{n}} \left((b-1)\binom{b+w-2}{n-2} + \binom{b+w-1}{n-1}\right) =$$

$$= \frac{b}{\binom{b+w}{n}} \left((b-1)\frac{(b+w-2)!}{(n-2)!(b+w-n)!} + \frac{(b+w-1)!}{(n-1)!(b+w-n)!}\right) =$$

$$= \frac{bn!(b+w-n)!}{(b+w)!} \left(\frac{(b-1)(n-1)(b+w-2)! + (b+w-1)!}{(n-1)!(b+w-n)!}\right) =$$

e

$$\sum_{k=0}^{n} k^2 \mathcal{P}(\xi = k) = \frac{bn\left((b-1)(n-1)+(b+w-1)\right)}{(b+w)!\left((b+w-2)!\right)} =$$
$$= \frac{bn\left(nb-n-b+1+b+w-1\right)}{(b+w)(b+w-1)} = \frac{bn\left(nb-n+w\right)}{(b+w)(b+w-1)}$$

da cui

$$\sigma^2 = \frac{bn\left(nb-n+w\right)}{(b+w)(b+w-1)} - \frac{b^2n^2}{(b+w)^2} =$$
$$= \frac{nb(nb-n+w)(b+w) - b^2n^2(b+w-1)}{(b+w)^2(b+w-1)} =$$
$$= \frac{nb(nb^2 - nb + bw + nbw - nw + w^2 - nb^2 - nbw + nb)}{(b+w)^2(b+w-1)} =$$
$$= \frac{nbw(b+w-n)}{(b+w)^2(b+w-1)}$$

Possiamo anche osservare che si ha

$$\frac{\binom{b}{k}\binom{w}{n-k}}{\binom{b+w}{n}} = \frac{b!}{k!(b-k)!}\frac{w!}{(n-k)!(w-(n-k))!}\frac{n!(b+w-n)!}{(b+w)!} =$$
$$= \binom{n}{k}\frac{b!}{(b-k)!}\frac{w!}{(w+k-n)!}\frac{(b+w-n)!}{(b+w)!} =$$
$$= \binom{n}{k}\frac{b(b-1)\cdots(b-k+1))w(w-1)\cdots(w+k-n+1)}{(b+w)(b+w-1)\cdots(b+w-n+1)} =$$
$$= \binom{n}{k}\frac{b}{b+w}\frac{b-1}{b+w-1}\cdots\frac{(b-k+1)}{b+w-k+1}\cdot$$
$$\frac{w}{b+w-k}\frac{w-1}{b+w-k-1}\cdots\frac{w-n+k+1}{b+w-n+1}$$

Ora se $b+w=N$, $\frac{b}{b+w}=p$ e $\frac{w}{b+w}=q$ si ha, dividendo numeratori e denominatori per $(b+w)$,

$$\frac{\binom{b}{k}\binom{w}{n-k}}{\binom{b+w}{n}} = \binom{n}{k}\frac{p}{1}\frac{p-\frac{1}{N}}{1-\frac{1}{N}}\cdots\frac{p-\frac{k-1}{N}}{1-\frac{k-1}{N}}\frac{q}{1-\frac{k}{N}}\frac{q-\frac{1}{N}}{1-\frac{k+1}{N}}\cdots\frac{q-\frac{n-k-1}{N}}{1-\frac{n-1}{N}}$$

e

$$\lim_{N\to+\infty}\frac{\binom{b}{k}\binom{w}{n-k}}{\binom{b+w}{n}} = \binom{n}{k}p^k q^{n-k}$$

Osserviamo anche che, con queste notazioni, la media e la varianza si esprimono come

$$\mu = np$$
$$\sigma^2 = npq\frac{N-n}{N-1}$$

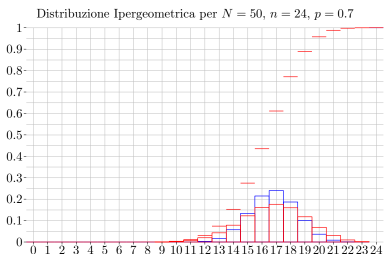


Figure 2.8: Confronto tra distribuzione Ipergeometrica e Binomiale.

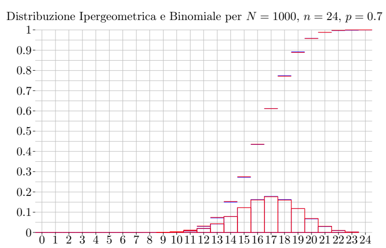


Figure 2.9: Confronto tra distribuzione Ipergeometrica e Binomiale.

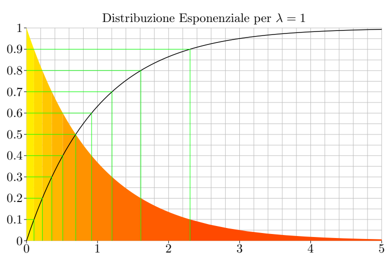


Figure 2.10: PDF e CDF di una variabile aleatoria Esponenziale.

Pertanto la media della distribuzione ipergeometrica è uguale alla media della distribuzione binomiale; per quanto concerne la varianza, possiamo vedere che il rapporto tra la varianza della ipergeometrica e la varianza della binomiale è dato da

$$\frac{N-n}{N-1} \to 1 \quad \text{per} \quad N \to +\infty$$

È pertanto evidente che per N grande la distribuzione ipergeometrica si riduce a quella binomiale.

In figura è riportata la PDF di una variabile aleatoria Binomiale e di una variabile aleatoria Ipergeometrica nel caso in cui $p = 0.7$, $q = 0.3$ $N = 7200$, $n = 24$.

2.4 *La distribuzione esponenziale*

Consideriamo ancora un centralino telefonico che in media riceve λ chiamate all'ora; abbiamo già visto che la variabile aleatoria che restituisce il numero il numero di chiamate in un'ora ha una distribuzione di Poisson di media e varianza λ.

Consideriamo ora la variabile aleatoria che restituisce il tempo che intercorre tra una chiamata e l'altra. A questo scopo conveniamo che

> $P_n(h)$ è la probabilità che si ricevano n chiamate in un intervallo di tempo di h ore.

dal momento che λh è la media di chiamate in un intervallo di h ore, usando la distribuzione di Poisson possiamo affermare che

$$P_0(h) = e^{-\lambda h}$$

Consideriamo ora la variabile aleatoria T che restituisce il tempo in cui avviene la prima chiamata a partire da 0.

Avremo che la probabilità che $T > t$, si calcola imponendo che in $[0, t]$ non si siano ricevute chiamate e quindi

$$\mathcal{P}(T > t) = P_0(t) = e^{-\lambda t}$$

Ne viene che

$$\mathcal{P}(T > t) = e^{-\lambda t} = \int_t^{+\infty} \varphi(t)dt$$

e possiamo trovare la PDF φ della distribuzione esponenziale semplicemente derivando rispetto a t

$$\varphi(t) = \lambda e^{-\lambda t}$$

Si verifica subito che media e varianza sono date da:

$$\mu = \int_0^{+\infty} \lambda t e^{-\lambda t} dt = -te^{-\lambda t}\Big|_0^{+\infty} + \int_0^{+\infty} e^{-\lambda t} dt = -\frac{e^{-\lambda t}}{\lambda}\Big|_0^{+\infty} = \frac{1}{\lambda}$$

mentre

$$\begin{aligned} \int_0^{+\infty} \lambda t^2 e^{-\lambda t} dt &= -t^2 e^{-\lambda t}\Big|_0^{+\infty} + \int_0^{+\infty} 2te^{-\lambda t} dt = \\ &= 2\left(-\frac{te^{-\lambda t}}{\lambda}\Big|_0^{+\infty} - \int_0^{+\infty} \frac{e^{-\lambda t}}{\lambda} dt\right) = 2\left(-\frac{e^{-\lambda t}}{\lambda^2}\Big|_0^{+\infty}\right) = \frac{2}{\lambda^2} \end{aligned}$$

per cui

$$\sigma = \frac{2}{\lambda^2} - \mu^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

2.5 *La distribuzione γ.*

La distribuzione γ è definita, per $\alpha, \beta > 0$ da

$$\boxed{\varphi(t) = \begin{cases} \frac{t^{\alpha-1}e^{-t/\beta}}{\beta^\alpha \Gamma(\alpha)} & t > 0 \\ 0 & \text{altrimenti} \end{cases}}$$

La presenza dei due parametri α e β consente di adattare la distribuzione ai dati che si desidera rappresentare. In particolare il parametro α descrive la forma della distribuzione mentre β è semplicemente un fattore di scala.

Media, varianza e generatrice dei momenti sono date da

$$\begin{aligned} \mu &= \alpha\beta \\ \sigma^2 &= \alpha\beta^2 \\ M_\gamma(t) &= (1-\beta t)^{-\alpha} \text{per } t < \frac{1}{\beta} \end{aligned}$$

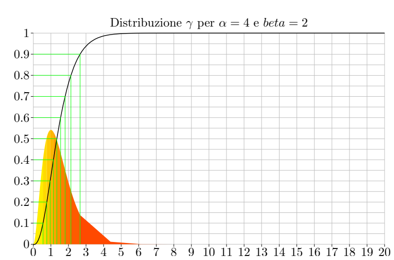


Figure 2.11: PDF e CDF di una variabile aleatoria γ.

Infatti

$$\begin{aligned} \mu = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^{+\infty} t^\alpha e^{-t/\beta} dt &= \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^{+\infty} (s\beta)^\alpha e^{-s} \beta ds = \\ &= \frac{\beta^{\alpha+1}\Gamma(\alpha+1)}{\beta^\alpha \Gamma(\alpha)} = \alpha\beta \end{aligned}$$

mentre

$$\frac{1}{\beta^\alpha\Gamma(\alpha)}\int_0^{+\infty} t^{\alpha+1}e^{-t/\beta}dt = \frac{1}{\beta^\alpha\Gamma(\alpha)}\int_0^{+\infty}(s\beta)^{\alpha+1}e^{-s}\beta ds =$$
$$= \frac{\beta^{\alpha+2}\Gamma(\alpha+2)}{\beta^\alpha\Gamma(\alpha)} = \alpha\beta^2(\alpha+1)$$

e

$$\sigma^2 = \beta^2\alpha^2 + \alpha\beta^2 - \beta^2\alpha^2 = \alpha\beta^2$$

Inoltre

$$M_\gamma(t) = \frac{1}{\beta^\alpha\Gamma(\alpha)}\int_0^{+\infty} e^{tx}x^{\alpha-1}e^{-x/\beta}dt =$$
$$= \frac{1}{\beta^\alpha\Gamma(\alpha)}\int_0^{+\infty}\left(\frac{1}{\left(\frac{1}{\beta}-t\right)}\right)^{\alpha-1} e^{-s}\frac{1}{\frac{1}{\beta}-t}ds =$$
$$= \frac{\beta^\alpha\Gamma(\alpha)}{\beta^\alpha\Gamma(\alpha)\left(\frac{1}{\beta}-t\right)^\alpha} = \frac{1}{\left(\frac{1}{\beta}-t\right)^\alpha}$$

Ovviamente il precedente integrale converge solo per $t < \frac{1}{\beta}$.

La moda infine è $(\alpha-1)\beta$ per $\alpha > 1$, come si deduce immediatamente calcolando il punto di massimo della funzione $t^{\alpha-1}e^{-t/\beta}$ la cui derivata

$$(\alpha-1)t^{\alpha-2}e^{-t/\beta} - \frac{1}{\beta}t^{\alpha-1}e^{-t/\beta} = t^{\alpha-2}e^{-t/\beta}\left((\alpha-1)-\frac{1}{\beta}\right) = 0$$

per

$$t = (\alpha-1)\beta$$

nel caso in cui $\alpha > 1$

Nel caso in cui $\alpha = k \in \mathbb{N}$ e $\beta = 1$ la distribuzione γ definisce una variabile aleatoria che è somma di k variabili aleatorie esponenziali di media $\lambda = \frac{1}{\beta}$. Infatti se consideriamo due variabili aleatorie con densità esponenziale di media λ la loro somma avrà una densità di probabilità definita dalla convoluzione

$$\psi(t) = \int_{-\infty}^{+\infty}\lambda e^{-\lambda s}\lambda e^{-\lambda(t-s)}ds = \int_0^t \lambda e^{-\lambda s}\lambda e^{-\lambda(t-s)}ds =$$
$$\lambda^2 e^{-\lambda t}\int_0^t ds = t\lambda^2 e^{-\lambda t}$$

Inoltre se sommiamo due variabili aleatorie con densità nulla prima di 0 e uguale a

$$\frac{t^{k-1}\lambda^k e^{-\lambda t}}{(k-1)!} \qquad \text{e} \qquad \lambda e^{-\lambda t}$$

altrove, otterremo una nuova variabile aleatoria la cui PDF si ottiene per convoluzione nella forma

$$\int_{-\infty}^{+\infty} \frac{1}{(k-1)!} s^{k-1} \lambda^k e^{-\lambda s} \lambda e^{-\lambda(t-s)} ds =$$
$$= \frac{1}{(k-1)!} \lambda^{k+1} e^{-\lambda t} \int_0^t s^{k-1} ds =$$
$$\frac{1}{k!} \lambda^{k+1} t^k e^{-\lambda t}$$

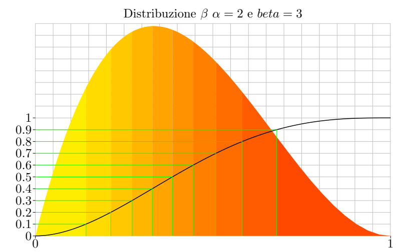


Figure 2.12: PDF e CDF di una variabile aleatoria β.

2.6 *La distribuzione β.*

La distribuzione β è definita da

$$\varphi(t) = \begin{cases} \frac{t^{\alpha-1}(1-t)^{\beta-1}}{B(\alpha,\beta)} & 0 < t < 1 \\ 0 & \text{altrimenti} \end{cases} = \tag{2.1}$$

$$= \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} t^{\alpha-1}(1-t)^{\beta-1} & 0 < t < 1 \\ 0 & \text{altrimenti} \end{cases} \tag{2.2}$$

Anche in questo caso si tratta di una distribuzione che, in virtù dei due parametri da cui dipende, ben si presta a descrivere una variabile aleatoria di cui si conoscano le caratteristiche.

I momenti della distribuzione β rispetto all'origine si calcolano facilmente non appena si ricordi che

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Si ha infatti

$$\mu_k = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^{+\infty} t^{\alpha-1+k}(1-t)^{\beta-1} dt =$$
$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+k)\Gamma(\beta)}{\Gamma(\alpha+\beta+k)} = \prod_{n=0}^{k-1} \frac{\alpha+n}{\alpha+\beta+n}$$

$\mu_1 = \mu$ fornisce la media e $\mu_2 - \mu_1^2 = \sigma^2$ fornisce la varianza che sono quindi date da:

$$\mu = \frac{\alpha}{\alpha+\beta}$$
$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}$$

La funzione generatrice dei momenti si può poi ricavare ricordando che i momenti sono i coefficienti del suo sviluppo di McLaurin.

$$M_\beta(t) = 1 + \sum_{k=1}^{+\infty} \left(\prod_{n=0}^{k-1} \frac{\alpha+n}{\alpha+\beta+n} \right) \frac{t^k}{k!}$$

Cercando il punto di massimo di $t^{\alpha-1}(1-t)^{\beta-1}$ si vede poi che la moda è

$$\frac{\alpha-1}{\alpha+\beta-2}$$

.

La distribuzione β, per $\alpha = n+1, \beta = m+1 \in \mathbb{N}$ diventa

$$\varphi(t) = \begin{cases} \frac{(n+m+)!}{n!\ m!} t^n (1-t)^m & 0 < t < 1 \\ 0 & \text{altrimenti} \end{cases}$$

e rappresenta la probabilità di aver ottenuto in una serie di $n+m$ prove bernoulliane ripetute, n successi ed m insuccessi, posto che sia t la probabilità di successo della prova bernoulliana stessa.

Possiamo dedurre da questa osservazione che il valore di t per cui questa probabilità è massima è $t = \frac{n}{n+m}$; infatti

$$\begin{aligned} (t^n(1-t)^m)' &= nt^{n-1}(1-t)^m + mt^n(1-t)^{m-1} = \\ &= (mt + (1-t)n)\left(t^{n-1}(1-t)^{m-1}\right) = 0 \iff t = \frac{n}{n+m} \end{aligned}$$

Possiamo quindi dire che in presenza di una sequenza di esiti bernoulliani con n successi ed m insuccessi è ragionevole supporre uguale a $\frac{n}{n+m}$ la probabilità di successo del singolo evento.

2.7 *La distribuzione Normale di Gauss*

La legge di distribuzione di probabilità normale è nota come distribuzione Gaussiana anche se Gauss si riferisce ad essa in una sua pubblicazione solo nel 1809.

In precedenza, nel 1733, De Moivre aveva pubblicato una derivazione della legge normale come limite di una distribuzione binomiale ed anche Laplace la conosceva già almeno dal 1774.

Gauss invece arrivò a considerare la distribuzione normale studiando il problema di stimare un parametro noto un certo numero di sue osservazioni. Per questo scopo applicò quello che si chiama oggi principio di massima verosimiglianza.

Una derivazione della legge normale molto interessante è dovuta ad Herschel , 1850, che la dedusse studiando la distribuzione a due dimensioni degli errori di misurazione della posizione di una stella. In seguito, nel 1860 James Clerk Maxwell estese le argomentazioni di

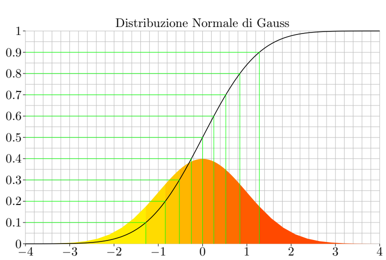


Figure 2.13: PDF e CDF di una variabile aleatoria Normale (di Gauss).

Herschel al caso tridimensionale, studiando la distribuzione di probabilità della velocità delle molecole in un gas.

Infine nel 1941 un ingegnere elettrico, Vernon D.Landon, studiando il rumore associato al voltaggio in circuito elettrico osservò che era distribuito con densita' invarianti rispetto all'intensità del disturbo e provò che questa caratteristica identifica la distribuzione normale.

La varietà e la diversità dei problemi che conducono alla distribuzione normale giustificano quindi pienamente il ruolo centrale che questa distribuzione occupa.

2.7.1 *La derivazione della distribuzione Normale di De Moivre*

Sia ξ_n la variabile aleatoria binomiale definita dalla somma di n variabili Bernoulliane ξ relative ad una prova con probabilità di successo p e probabilità di insuccesso $q = 1 - p$.

$$\xi_n = \xi + \xi + \xi + ... + \xi$$

e sia ζ_n la corrispondente variabile aleatoria normalizzata.

Si può dimostrare che se $n \to +\infty$ la distribuzione di probabilità di ζ_n tende ad una distribuzione normale .

Si dimostra in tal modo un caso particolare del teorema del limite centrale che verrà trattato più avanti

2.7.2 *La derivazione della distribuzione normale di Herschel-Maxwell*

Una stella si individua mediante la sua longitudine, misurata con un errore ξ, e la sua declinazione, misurata con un errore η; Herschel (1850) postulò che ξ e η sono variabili aleatorie indipendenti ed hanno la stessa PDF f. Allora la distribuzione di probabilità della variabile aleatoria (ξ, η) è della forma

$$\phi(x, y) = f(x)f(y)$$

Inoltre postulò che ϕ, espressa usando le coordinate polari (ρ, θ), fosse indipendente da θ, per cui

$$\phi(x, y) = f(x)f(y) = g(x^2 + y^2)$$

Ne dedusse infine che ϕ doveva allora essere quella che ora chiamiamo densità di probabilità gaussiana.

Maxwell (1860) usò argomentazioni sostanzialmente identiche nello studio della cinetica dei gas estendendo l'idea di Herschel al caso tridimensionale.

Infatti, ponendo $x = 0$ e $f(0) = \alpha$ si ricava $f(x)\alpha = g(x^2)$ da cui

$$\frac{g(x^2)}{\alpha}\frac{g(y^2)}{\alpha} = g(x^2 + y^2)$$

f additiva e continua è lineare, infatti in tal caso si ha

$$f(nx) = nf(x)$$

e di conseguenza

$$f\left(\frac{x}{m}\right) = \frac{1}{m}f(x)$$

per ogni n intero, e per ogni m naturale. Se ne deduce che

$$f\left(\frac{n}{m}x\right) = \frac{n}{m}f(x)$$

da cui

$$f(qx) = qf(x)$$

per ogni q razionale. La continuità garantisce poi che

$$f(\alpha x) = \alpha f(x)$$

per ogni x reale e l'omogeneità di f che essendo già additiva risulta lineare.

dividendo per α^2

$$\frac{g(x^2)}{\alpha^2}\frac{g(y^2)}{\alpha^2} = \frac{g(x^2+y^2)}{\alpha^2}$$

e

$$\ln\left(\frac{g(x^2)}{\alpha^2}\right) + \ln\left(\frac{g(y^2)}{\alpha^2}\right) = \ln\left(\frac{g(x^2+y^2)}{\alpha^2}\right)$$

Pertanto la funzione $z \mapsto \ln\left(\frac{g(z)}{\alpha^2}\right)$ è lineare e quindi

$$\ln\left(\frac{g(z)}{\alpha^2}\right) = kz \quad \text{e} \quad \ln\left(\frac{g(x^2)}{\alpha^2}\right) = kx^2$$

Ne deduciamo che

$$g(x^2) = h^2 e^{kx^2}$$

e le costanti h e k possono essere determinate in modo da aversi una distribuzione di probabilità. Innanzi tutto si vede che deve essere $h > 0$ e $k < 0$, imponendo poi che la distribuzione abbia media μ e varianza σ otteniamo

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

2.7.3 *La derivazione della distribuzione normale di Gauss*

Gauss considerò il problema di stimare un parametro θ note $n+1$ sue osservazioni $x_0, x_1, \ldots, x_n$ ed usò allo scopo il principio di massima verosimiglianza.

Se indichiamo con $f(x_i, \theta)$ la probabilità di ottenere la misura x_i condizionata al fatto che il parametro cercato è θ, la probabilità di avere ottenuto le osservazioni x_i, che supponiamo indipendenti, è data da

$$P(\theta) = \prod f(x_i, \theta)$$

Cerchiamo di determinare f in modo che $P(\theta)$ sia massima in corrispondenza del valor medio $\hat{\theta}$ delle osservazioni. L'argomento di massimo non cambia se consideriamo $\ln(P(\theta))$ in luogo di $P(\theta)$ e quindi θ deve soddisfare la condizione

$$\sum_0^n \frac{\partial}{\partial\theta} \ln(f(x_i, \theta)) = 0$$

Se poniamo

$$\ln(f(x_i, \theta)) = g(\theta - x_i) = g(u)$$

se supponiamo cioe' che $f(x_i, \theta)$ dipenda solo dall'errore commesso nel considerare θ invece di x_i, dovrà risultare

$$\sum_0^n g'(\theta - x_i) = 0 \tag{2.3}$$

Se vogliamo che il massimo sia assunto per

$$\theta = \hat{\theta} = \frac{1}{n+1}\sum_0^n x_i \tag{2.4}$$

dovrà aversi

$$\sum_0^n g'(\hat{\theta} - x_i) = 0 \tag{2.5}$$

Le equazioni 2.5, 2.4 sono, in generale, incompatibili, anzi possiamo subito osservare che qualora la distribuzione degli errori di misurazione sia uniforme si avrebbe $f(x_i, \theta)$ costante per cui tale risulterebbe anche $P(\theta)$ e la 2.5 perderebbe di significato attribuendo a qualunque valore di θ la stessa affidabilità.

Ora, se consideriamo il caso in cui una sola delle osservazioni differisce dalla media, cioè se poniamo

$$x_0 = (n+1)u, \qquad x_1 = x_2 = \cdots = x_n = 0$$

per modo che

$$\hat{\theta} = u, \qquad \hat{\theta} - x_0 = -nu, \qquad \hat{\theta} - x_i = u - 0, i = 1, \cdots n$$

sostituendo in 2.5 otteniamo

$$g'(-nu) + ng'(u) = 0 \tag{2.6}$$

Per $n = 1$ la 2.6 fornisce

$$g'(-u) = -g'(u)$$

e ne viene che g' deve essere antisimmetrica inoltre la 2.6, garantisce che la g', supposta continua, è lineare.

La linearità permette di concludere che

$$g'(u) = au, \qquad g(u) = \frac{1}{2}au^2 + b$$

Infine usando la definizione di g e tenendo conto che la distribuzione di probabilità deve essere normalizzata, otteniamo che $a < 0$ ed anche

$$f(x, \theta) = \sqrt{\frac{\alpha}{2\pi}} e^{-\frac{1}{2}\alpha(x-\theta)^2} \tag{2.7}$$

Nel caso in cui la 2.7 valga la funzione che definisce la verosimiglianza diventa

$$\ln(P(\theta)) = \sum_0^n \ln\left(\sqrt{\frac{\alpha}{2\pi}} e^{-\frac{1}{2}\alpha(x_i-\theta)^2}\right) =$$
$$= (n+1)\ln\left(\sqrt{\frac{\alpha}{2\pi}}\right) + \sum_0^n -\frac{1}{2}\alpha(x_i - \theta)^2$$

Ne segue che

$$(\ln(P(\theta)))' = \sum_0^n \alpha(x_i - \theta) = \sum_0^n \alpha x_i - (n+1)\theta$$

Che evidentemente assume massimo proprio per $\theta = \hat{\theta}$ ed inoltre il punto di massimo è unico (si tratta di un paraboloide concavo). Resta libero il parametro α che definisce diverse forme della distribuzione.

2.7.4 *La derivazione della distribuzione Normale di Landon*

Landon studiò la distribuzione di probabilità della variabile aleatoria ξ che rappresenta il rumore da cui è affetto il voltaggio osservato in un circuito elettrico. Egli osservò che tale distribuzione sembrava non cambiare forma al variare della deviazione standard σ del rumore.

Sia quindi $p(x, \sigma)$ la probabilità che il rumore valga x nel caso in cui la sua deviazione standard sia σ.

Supponiamo che alla variabile aleatoria ξ si sommi una quantità aleatoria $\Delta\xi$ indipendente da ξ la cui distribuzione si chiami q. Avremo che

$$f(x) = \int_{-\infty}^{+\infty} p(x-\epsilon, \sigma^2) q(\epsilon) d\epsilon$$

è la distribuzione di probabilità della variabile aleatoria $\xi + \Delta\xi$

Usando lo sviluppo di Taylor di ordine 2 di $p(x, \sigma)$ otteniamo

$$f(x) = \\ = \int_{-\infty}^{+\infty} q(\epsilon) \left(p(x, \sigma^2) - \epsilon \frac{\partial p(x, \sigma^2)}{\partial x} + \epsilon^2 \frac{\partial^2 p(x, \sigma)}{2\partial x^2} + R \right) d\epsilon$$

e

$$f(x) = p(x, \sigma^2) - \mu(q) \frac{\partial p(x, \sigma^2)}{\partial x} + Var(q) \frac{1}{2} \frac{\partial p(x, \sigma^2)}{\partial x^2} + R$$

Pertanto supponendo che la media $\mu(q)$ sia nulla e trascurando i momenti di ordine superiore al secondo, si ha

$$f(x) = p(x, \sigma^2) + Var(q) \frac{1}{2} \frac{\partial^2 p(x, \sigma^2)}{\partial x^2} + R$$

D'altro canto, poichè si suppone che la forma della distribuzione di probabilità del rumore non cambi con la varianza e dal momento che $Var(\xi + \Delta\xi) = \sigma^2 + Var(\Delta\xi)$, deve essere

$$f(x) = p(x, \sigma^2 + Var(q)) = p(x, \sigma^2) + Var(q)\frac{\partial p(x,\sigma^2)}{\partial \sigma} + R$$

Ove si sia usato lo sviluppo di Taylor di $p(x, \cdot)$.

Possiamo quindi dedurre confrontando i due sviluppi che

$$\frac{\partial p(x,\sigma^2)}{\partial \sigma} = \frac{1}{2}\frac{\partial^2 p(x,\sigma^2)}{\partial x^2}$$

che è una equazione di diffusione che , con condizioni iniziali $p(x, 0) = \delta(x)$ fornisce come soluzione

$$p(x, \sigma^2) = N(x, 0, \sigma)$$

2.8 Combinazione lineare di variabili gaussiane

Siano ξ_1 e ξ_2 variabili aleatorie gaussiane di media 0 e varianza 1 e consideriamo $\xi = \alpha\xi_1 + \beta\xi_2$ con $\alpha^2 + \beta^2 = 1$, allora ξ è una variabile aleatoria normale standard.

Avremo che la PDF di ξ_1 e di ξ_2 è data da

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$$

e si può affermare che $\alpha\xi_1$ e $\beta\xi_2$ hanno rispettivamente, le seguenti PDF:

$$\frac{1}{\sqrt{2\pi}|\alpha|} e^{-\frac{u^2}{2\alpha^2}} \quad , \quad \frac{1}{\sqrt{2\pi}|\beta|} e^{-\frac{u^2}{2\beta^2}}$$

Ne viene che la PDF ϕ della somma ξ sarà data da

$$\phi(v) = \frac{1}{\sqrt{2\pi}|\alpha|}\frac{1}{\sqrt{2\pi}|\beta|}\int_{-\infty}^{+\infty} e^{-\frac{u^2}{2\alpha^2}} e^{-\frac{(v-u)^2}{2\beta^2}} du$$

e si calcola

$$\begin{aligned}
\phi(v) &= \frac{1}{2\pi|\alpha\beta|}\int_{-\infty}^{+\infty} e^{-\frac{u^2}{2\alpha^2} - \frac{v^2}{2\beta^2} - \frac{u^2}{2\beta^2} + \frac{uv}{\beta^2}} du = \\
&= \frac{1}{2\pi|\alpha\beta|} e^{-\frac{v^2}{2\beta^2}} \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2\alpha^2} - \frac{u^2}{2\beta^2} + \frac{uv}{\beta^2}} du = \\
&= \frac{1}{2\pi|\alpha\beta|} e^{-\frac{v^2}{2\beta^2}} \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2\alpha^2\beta^2} + \frac{uv}{\beta^2}} du = \\
&= \frac{1}{2\pi|\alpha\beta|} e^{-\frac{v^2}{2\beta^2}} \int_{-\infty}^{+\infty} e^{-\left(\frac{u}{\sqrt{2}|\alpha\beta|} - \frac{v}{\sqrt{2}\beta^2}|\alpha\beta|\right)^2 + \frac{v^2}{2\beta^4}\alpha^2\beta^2} du = \\
&= \frac{1}{2\pi|\alpha\beta|} e^{-\frac{v^2}{2\beta^2} + \frac{v^2}{2\beta^4}\alpha^2\beta^2} \int_{-\infty}^{+\infty} e^{-\left(\frac{u}{\sqrt{2}|\alpha\beta|} - \frac{v}{\sqrt{2}\beta^2}|\alpha\beta|\right)^2} du
\end{aligned}$$

Ora, tenendo conto che

$$-\frac{1}{2\beta^2}+\frac{1}{2\beta^4}\alpha^2\beta^2=-\frac{1}{2}$$

e calcolato l'integrale ponendo

$$t=\frac{u}{\sqrt{2}|\alpha\beta|}-\frac{v}{\sqrt{2}\beta^2}|\alpha\beta|$$

si conclude

$$\phi(v)=\frac{1}{2\pi|\alpha\beta|}e^{-\frac{v^2}{2}}\sqrt{2\pi}|\alpha\beta|=\frac{1}{\sqrt{2\pi}}e^{-\frac{v^2}{2}}$$

e ξ è a sua volta normale e standard.

Chiaramente qualora $\alpha^2+\beta^2\neq 1$ possiamo considerare che

$$\alpha\xi_1+\beta\xi_2=\sqrt{\alpha^2+\beta^2}\left(\frac{\alpha}{\sqrt{\alpha^2+\beta^2}}\xi_1+\frac{\beta}{\sqrt{\alpha^2+\beta^2}}\xi_2\right)$$

inoltre, se η_1 e η_2 sono gaussiane di media μ_i e varianza σ_i^2 possiamo scrivere che

$$\begin{aligned}\alpha\eta_1+\beta\eta_2&=\alpha\mu_1+\alpha\sigma_1\xi_1+\beta\mu_2+\beta\sigma_2\xi_2=\\&=\alpha\mu_1+\beta\mu_2+(\alpha\sigma_1\xi_1+\beta\sigma_2\xi_2)=\\\alpha\mu_1+\beta\mu_2+\sqrt{\alpha^2\sigma_1^2+\beta^2\sigma_2^2}&\left(\frac{\alpha\sigma_1}{\sqrt{\alpha^2\sigma_1^2+\beta^2\sigma_2^2}}\xi_1+\frac{\beta\sigma_2}{\sqrt{\alpha^2\sigma_1^2+\beta^2\sigma_2^2}}\xi_2\right)=\\&\alpha\mu_1+\beta\mu_2+\sqrt{\alpha^2\sigma_1^2+\beta^2\sigma_2^2}\xi\end{aligned}$$

essendo ξ una gaussiana normale standard.

Ne segue che $\alpha\eta_1+\beta\eta_2$ ha una distribuzione gaussiana di media $\alpha\mu_1+\beta\mu_2$ e varianza $\alpha^2\sigma_1^2+\beta^2\sigma_2^2$

in quanto

$$\left(\frac{\alpha\sigma_1}{\sqrt{\alpha^2\sigma_1^2+\beta^2\sigma_2^2}}\xi_1+\frac{\beta\sigma_2}{\sqrt{\alpha^2\sigma_1^2+\beta^2\sigma_2^2}}\xi_2\right)$$

è una gaussiana standard.

2.9 *Le distribuzioni legate ai test statistici.*

2.9.1 *La distribuzione χ^2*

Si tratta della distribuzione di probabilità di una variabile aleatoria χ^2 che restituisce la somma dei quadrati di ν variabili aleatorie ξ_i indipendenti, aventi distribuzione gaussiano con media 0 e varianza 1 (distribuzioni normali standardizzate).

$$\chi^2=\xi_1^2+\xi_2^2+...+\xi_\nu^2$$

Per ricavare la PDF della distribuzione χ^2 a ν gradi di libertà è opportuno procedere come segue.

2.9.2 χ^2 *ad* 1 *grado di libertà*

Se ξ è una variabile aleatoria gaussiana standard e se $\eta = \xi^2$, la PDF φ di η è data da

$$\varphi(t) = \begin{cases} \frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{s}{2}}}{\sqrt{s}} & s \geq 0 \\ 0 & s < 0 \end{cases}$$

Infatti si ha

$$\begin{aligned} \mathcal{P}(\eta \leq \alpha) &= \mathcal{P}(\xi^2 \leq \alpha) = \mathcal{P}(-\sqrt{\alpha} \leq \xi \leq \sqrt{\alpha}) = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{\alpha}}^{\sqrt{\alpha}} e^{-\frac{t^2}{2}} dt = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{\alpha}} e^{-\frac{t^2}{2}} dt = \end{aligned}$$

posto $s = t^2$, per cui $ds = 2tdt$

$$= \frac{2}{\sqrt{2\pi}} \int_0^{\alpha} \frac{e^{-\frac{s}{2}}}{2\sqrt{s}} ds = \frac{1}{\sqrt{2\pi}} \int_0^{\alpha} \frac{e^{-\frac{s}{2}}}{\sqrt{s}} ds$$

2.9.3 χ^2 *a* 2 *gradi di libertà*

Se η_1, η_2 sono variabili aleatorie $\eta_i = \xi_i^2$ con ξ_i gaussiana standard e se $\eta = \eta_1 + \eta_2$, la PDF φ di η è data da

$$\varphi(t) = \begin{cases} \frac{1}{2} e^{-\frac{s}{2}} & s \geq 0 \\ 0 & s < 0 \end{cases}$$

Infatti, se chiamiamo φ_i la PDF di η_i, si ha

$$\begin{aligned} \varphi(s) &= \int_{-\infty}^{+\infty} \varphi_1(t)\varphi_2(s-t)dt = \\ &= \int_0^s \varphi_1(t)\varphi_2(s-t)dt = \frac{1}{2\pi} \int_0^s \frac{e^{-\frac{t}{2}} e^{-\frac{s-t}{2}}}{\sqrt{t}\sqrt{s-t}} dt = \\ &= \frac{1}{2\pi} e^{-\frac{s}{2}} \int_0^s \frac{1}{\sqrt{t}\sqrt{s-t}} dt \end{aligned}$$

Posto $\sqrt{t(s-t)} = ut$ si ha

$$t(s-t) = u^2t^2 \ , \ s - t = u^2 t \ , \ s = (1-u^2)t \ , \ t = \frac{s}{1+u^2} \ , \ dt = -\frac{2us}{(1+u^2)^2} du$$

per cui

$$\int \frac{1}{\sqrt{t(s-t)}}dt = \int -\frac{2us}{ut(1+u^2)^2}du =$$
$$= \int -\frac{2us}{u\frac{s}{1+u^2}(1+u^2)^2}du = -2\int \frac{1}{1+u^2}du =$$
$$= -2\arctan(u) = -2\arctan\sqrt{\frac{s-t}{t}}$$

Per cui

$$\int_0^s \frac{1}{\sqrt{t}\sqrt{s-t}}dt = -2\arctan\sqrt{\frac{s-t}{t}}\Big|_0^s = 2\frac{\pi}{2} = \pi$$

e

$$\varphi(s) = \frac{1}{2\pi}e^{-\frac{s}{2}}\pi = \frac{1}{2}e^{-\frac{s}{2}}$$

2.9.4 χ^2 *a* ν *gradi di libertà*

Possiamo ora provare che la variabile aleatoria

$$\eta = \eta_1 + \eta_2 + \cdot + \eta_\nu = \xi_1^2 + \xi_2^2 + \cdot + \xi_\nu^2$$

dove ξ_i sono variabili aleatorie gaussiane normalizzate, ha una PDF definita da

$$\varphi_\nu(u) = \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}u^{\frac{\nu}{2}-1}e^{-\frac{u}{2}}$$

Le verifiche fatte precedentemente consentono di affermare che la precedente affermazione è vera per $\nu = 1$ e per $\nu = 2$. Pertanto per verificare la medesima è sempre vera sarà sufficiente provare che, supposta vera per ν è vera anche per $\nu + 2$.

Si ha

$$\varphi_{\nu+2}(u) = \varphi_\nu(u) * \varphi_2(u) = \int_0^u \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}(u-t)^{\frac{\nu}{2}-1}e^{-\frac{u-t}{2}}\frac{1}{2}e^{-\frac{t}{2}}dt =$$
$$= \frac{1}{2^{\frac{\nu}{2}+1}\Gamma\left(\frac{\nu}{2}\right)}e^{-\frac{u}{2}}\int_0^u (u-t)^{\frac{\nu}{2}-1}du = \frac{1}{2^{\frac{\nu}{2}+1}\Gamma\left(\frac{\nu}{2}\right)}e^{-\frac{u}{2}}\int_0^u (u-t)^{\frac{\nu}{2}-1}du =$$
$$= \frac{1}{2^{\frac{\nu}{2}+1}\Gamma\left(\frac{\nu}{2}\right)}e^{-\frac{u}{2}}\left[-\frac{(u-t)^{\frac{\nu}{2}}}{\frac{\nu}{2}}\right]_0^u = \frac{1}{2^{\frac{\nu}{2}+1}\Gamma\left(\frac{\nu}{2}\right)}e^{-\frac{u}{2}}\frac{(u)^{\frac{\nu}{2}}}{\frac{\nu}{2}} =$$
$$= \frac{1}{2^{\frac{\nu}{2}+1}\Gamma\left(\frac{\nu}{2}+1\right)}u^{\frac{\nu}{2}}e^{-\frac{u}{2}}$$

Media, varianza e generatrice dei momenti della variabile aleatoria χ^2 sono date da

$$\mu = \nu$$
$$\sigma^2 = 2\nu$$
$$M_\chi^2(t) = (1-2t)^{-\nu/2}$$

2.9.5 *La distribuzione T di Student.*

È la distribuzione di una variabile aleatoria T che restituisce il rapporto

$$T = \frac{\xi}{\sqrt{\frac{\eta}{\nu}}}$$

dove ξ è una variabile aleatoria con densità di probabilità gaussiana normale (media 0 e varianza 1) ed η è una variabile aleatoria con distribuzione χ^2 a ν gradi di libertà

Per ricavare la PDF di student cominciamo con l'osservare che:

2.9.6

Se ξ è una variabile aleatoria con densità χ^2 a ν gradi di libertà, la sua PDF sarà data da

$$\varphi_\nu(u) = \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} u^{\frac{\nu}{2}-1} e^{-\frac{u}{2}}$$

Sia η la variabile aleatoria definita da

$$\eta = \sqrt{\frac{\xi}{\nu}}$$

Avremo che

$$\mathcal{P}(\eta \leq \alpha) = \mathcal{P}\left(\sqrt{\frac{\xi}{\nu}} \leq \alpha\right) = \mathcal{P}(\xi \leq \nu\alpha^2) = \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} \int_0^{\nu\alpha^2} u^{\frac{\nu}{2}-1} e^{-\frac{u}{2}} du =$$

Posto $t = \sqrt{\frac{u}{\nu}}$ da cui $u = \nu t^2$ e $du = 2\nu t dt$

$$= \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} \int_0^{\alpha} t^{\nu-2} \nu^{\frac{\nu}{2}-1} e^{-\frac{\nu t^2}{2}} 2\nu t dt =$$

$$= \frac{1}{2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} \int_0^{\alpha} t^{\nu-1} \nu^{\frac{\nu}{2}} e^{-\frac{\nu t^2}{2}} dt$$

e possiamo concludere che la PDF di η è data da

$$\varphi(t) = \begin{cases} \frac{1}{2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} t^{\nu-1} \nu^{\frac{\nu}{2}} e^{-\frac{\nu t^2}{2}} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

2.9.7

Sia ora η_1 una variabile aleatoria gaussiana standard, la cui PDF è ovviamente data da

$$g(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$$

e sia $\eta_2 = \sqrt{\frac{\xi}{\nu}}$ dove ξ è una variabile aleatoria con densità χ^2 a ν gradi di libertà; per quanto detto in precedenza la PDF di η_2 è nulla prima di 0 ed è data da

$$g(t) = \frac{1}{2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} t^{\nu-1}\nu^{\frac{\nu}{2}} e^{-\frac{\nu t^2}{2}}$$

per $t \geq 0$.

La variabile aleatoria T di student è definita mediante la

$$T = \frac{\eta_1}{\eta_2}$$

e possiamo ricavarne la PDF φ ricordando che si tratta del quoziente di due variabili aleatorie di cui conosciamo la densità di probabilità. Avremo:

$$\varphi(t) = \int_0^{+\infty} x \frac{1}{2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} x^{\nu-1}\nu^{\frac{\nu}{2}} e^{-\frac{\nu x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2 x^2}{2}} dx =$$

$$= \frac{\nu^{\frac{\nu}{2}}}{\sqrt{2\pi}2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} \int_0^{+\infty} x^{\nu} e^{-\frac{x^2}{2}(\nu+t^2)} dx =$$

posto $\frac{x^2}{2}(\nu + t^2) = s$ si ha $x = \sqrt{\frac{2s}{\nu+t^2}}$ e $dx = \frac{\sqrt{2}}{\sqrt{\nu+t^2}2\sqrt{s}}$

$$= \frac{\nu^{\frac{\nu}{2}}}{\sqrt{2\pi}2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} \int_0^{+\infty} \frac{2^{\frac{\nu}{2}} s^{\frac{\nu}{2}}}{(\nu+t^2)^{\frac{\nu}{2}}} e^{-s} \frac{\sqrt{2}}{\sqrt{\nu+t^2}} \frac{1}{2\sqrt{s}} ds =$$

$$= \frac{\nu^{\frac{\nu}{2}} 2^{\frac{\nu}{2}-\frac{1}{2}}}{\sqrt{2\pi}2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{(\nu+t^2)^{\frac{\nu+1}{2}}} \int_0^{+\infty} s^{\frac{\nu}{2}-\frac{1}{2}} e^{-s} ds =$$

$$= \frac{\nu^{\frac{\nu}{2}}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{(\nu)^{\frac{\nu+1}{2}}(1+\frac{t^2}{\nu})^{\frac{\nu+1}{2}}} \int_0^{+\infty} s^{\frac{\nu+1}{2}-1} e^{-s} ds =$$

$$= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)(1+\frac{t^2}{\nu})^{\frac{\nu+1}{2}}}$$

Pertanto la funzione di distribuzione della variabile aleatoria T a ν gradi di libertà è data da

$$\boxed{\varphi(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}}$$

La media e la varianza di T risultano essere

$$\mu = 0$$
$$\sigma^2 = \frac{\nu}{\nu-2}$$

2.9.8 La distribuzione F di Fisher.

È la distribuzione di una variabile aleatoria F che restituisce il rapporto

$$F = \frac{\eta/\mu}{\xi/\nu}$$

dove η ed ξ sono variabili aleatorie con distribuzione χ^2 a μ e ν gradi di libertà, rispettivamente.

Per ricavare la PDF di F ricordiamo che la PDF g di η/μ è data da

$$g(u) = \frac{\mu}{2^{\frac{\mu}{2}}\Gamma\left(\frac{\mu}{2}\right)}(\mu u)^{\frac{\mu}{2}-1}e^{-\frac{\mu u}{2}}$$

mentre la PDF f di ξ/ν è data da

$$f(u) = \frac{\nu}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}(\nu u)^{\frac{\nu}{2}-1}e^{-\frac{\nu u}{2}}$$

Possiamo quindi ricavare la PDF φ di $F = \frac{\eta/\mu}{\xi/\nu}$ usando quanto conosciamo sul rapporto di due variabili aleatorie. Avremo

$$\varphi(t) = \int_0^{+\infty} x\frac{\nu}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}\nu^{\frac{\nu}{2}-1}x^{\frac{\nu}{2}-1}e^{-\frac{\nu x}{2}}\frac{\mu}{2^{\frac{\mu}{2}}\Gamma\left(\frac{\mu}{2}\right)}\mu^{\frac{\mu}{2}-1}(tx)^{\frac{\mu}{2}-1}e^{-\frac{\mu t x}{2}}dx =$$

$$= \frac{\nu^{\frac{\nu}{2}}\mu^{\frac{\mu}{2}}}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)2^{\frac{\mu}{2}}\Gamma\left(\frac{\mu}{2}\right)}\int_0^{+\infty} x^{\frac{\nu}{2}}e^{-\frac{\nu x}{2}}t^{\frac{\mu}{2}-1}x^{\frac{\mu}{2}-1}e^{-\frac{\mu t x}{2}}dx =$$

$$= t^{\frac{\mu}{2}-1}C_{\nu,\mu}\int_0^{+\infty} x^{\frac{\nu+\mu}{2}-1}e^{-\frac{x}{2}(\nu+t\mu)}dx =$$

posto $\frac{x}{2}(\nu+t\mu) = s$ si ha $dx = \frac{2}{\nu+t\mu}ds$

$$= t^{\frac{\mu}{2}-1}C_{\nu,\mu}\int_0^{+\infty}\left(\frac{2s}{\nu+t\mu}\right)^{\frac{\nu+\mu}{2}-1}e^{-s}\frac{2}{\nu+t\mu}ds =$$

$$= \frac{t^{\frac{\mu}{2}-1}C_{\nu,\mu}2^{\frac{\nu+\mu}{2}}}{(\nu+t\mu)^{\frac{\nu+\mu}{2}}}\int_0^{+\infty} s^{\frac{\nu+\mu}{2}-1}e^{-s}ds =$$

$$= \frac{\nu^{\frac{\nu}{2}}\mu^{\frac{\mu}{2}}\Gamma\left(\frac{\nu+\mu}{2}\right)2^{\frac{\nu+\mu}{2}}}{2^{\frac{\nu}{2}}2^{\frac{\mu}{2}}\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{\mu}{2}\right)}\frac{t^{\frac{\mu}{2}-1}}{(\nu+t\mu)^{\frac{\nu+\mu}{2}}} =$$

$$= \frac{\nu^{\frac{\nu}{2}}\mu^{\frac{\mu}{2}}\Gamma\left(\frac{\nu+\mu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{\mu}{2}\right)}t^{\frac{\mu}{2}-1}(\nu+t\mu)^{-\frac{\nu+\mu}{2}}$$

Pertanto la funzione di distribuzione della variabile aleatoria F è data da

$$\boxed{\varphi(t) = \begin{cases} \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)}\nu_1^{\nu_1/2}\nu_2^{\nu_2/2}t^{\nu_2/2-1}(\nu_2+\nu_1 t)^{-(\nu_1+\nu_2)/2} & t > 0 \\ 0 & \text{altrimenti} \end{cases}}$$

la media e varianza risultano essere

$$\mu = \frac{\nu_2}{\nu_2 - 2}$$
$$\sigma^2 = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 4)(\nu_2 - 2)^2}$$

2.10 *Variabili casuali con distribuzione assegnata.*

Qualora sia necessario utilizzare dati generati casualmente con funzione distribuzione di probabilità fissata, possiamo procedere come segue.

Sia ϕ la distribuzione che si vuole considerare e sia

$$F(x) = \int_{-\infty}^{x} \phi(t)dt$$

la funzione di distribuzione cumulativa; allora si ha

$$\mathcal{P}(a \leq \xi \leq b) = \int_{a}^{b} \phi(t)dt = F(b) - F(a)$$

e quindi

$$\mathcal{P}(F^{-1}(a) \leq \xi \leq F^{-1}(b)) = b - a$$

e

$$\mathcal{P}(a \leq F(\xi) \leq b) = b - a$$

Pertanto $F(\xi)$ ha una distribuzione uniforme e quindi poichè

$$\xi = F^{-1}(F(\xi))$$

possiamo generare valori distribuiti con densità di probabilità ϕ, considerando valori generati con densità uniforme ed applicando a tali valori F^{-1}.

Tale procedimento non è tuttavia applicabile, ad esempio, per determinare valori distribuiti con densità gaussiana in quanto non è possibile determinare esplicitamente F^{-1} nel caso in cui

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}dt$$

In tal caso, che peraltro è di rilevante importanza possiamo osservare che se ξ e η sono variabili aleatorie indipendenti con distribuzione gaussiana normale, allora (ξ, η) è una variabile aleatoria la cui funzione distribuzione di probabilità è

$$\frac{1}{2\pi} e^{-(t^2+s^2)/2}$$

Pertanto

$$\mathcal{P}((\xi,\eta)\in A) = \frac{1}{2\pi}\iint\limits_A e^{-(t^2+s^2)/2}dtds = \frac{1}{2\pi}\iint\limits_B \rho e^{-(\rho^2)/2}d\rho d\theta$$

dove B ed A sono l'uno il trasformato dell'altro rispetto al cambio di variabili in coordinate polari.

Ne viene che possiamo identificare due nuove variabili (R,Θ) la cui densità di probabilità è data da

$$\frac{1}{2\pi}\rho e^{-(\rho^2)/2} = (\frac{1}{2\pi})(\rho e^{-(\rho^2)/2}) = f(\theta)g(\rho)$$

dove

$$f(\theta) = \frac{1}{2\pi}$$
$$g(\rho) = \rho e^{-(\rho^2)/2}$$

Quindi per quanto visto in precedenza, per generare valori casuali di Θ e di R possiamo utilizzare valori uniformemente distribuiti θ ed r e applicare a tali valori le funzioni F^{-1} e G^{-1}, rispettivamente, dove

$$F(t) = \int_0^t f(s)ds = \frac{t}{2\pi} \quad , \quad G(t) = \int_0^t g(s)ds = 1 - e^{-t^2/2}$$

Si ha allora

$$F^{-1}(s) = 2\pi s \quad , \quad G^{-1}(s) = \sqrt{-2\ln(1-s)}$$

e le variabili

$$\xi = \sqrt{-2\ln(1-t)}\cos(2\pi s) \quad , \quad \eta = \sqrt{-2\ln(1-t)}\sin(2\pi s)$$

dove s e t sono distribuite uniformemente, risultano distribuite con densità gaussiana di media 0 e di varianza 1.